Serial No. 10/516,578
Group No. 1648
Confirmation No. 5513

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 December 2003 (11.12.2003)

PCT

(10) International Publication Number
WO 2003/102219 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number:
PCT/US2003/017577

(22) International Filing Date: 4 June 2003 (04.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/386,064 4 June 2002 (04.06.2002) US
60/458,070 27 March 2003 (27.03.2003) US

(71) Applicants *(for all designated States except US)*: **PURDUE RESEARCH FOUNDATION** [US/US]; 1291 Cumberland Avenue, West Lafayette, IN 47906 (US). **CENTERS FOR DISEASE CONTROL AND PREVENTION** [US/US]; 1600 Clifton Road Northeast, Atlanta, GA 30329 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SANDERS, David, A.** [US/US]; 324 Jefferson Drive, West Lafayette, IN 47906 (US). **JEFFERS, Scott, A.** [US/US]; 1945 Indian Trail Drive, West Lafayette, IN 47906 (US). **SANCHEZ, Anthony** [US/US]; 1717 Red Fox Run, Lilburn, GA 30047 (US).

(74) Agent: **SANDBERG, Victoria, A.**; Mueting, Raasch & Gebhardt, P.A., P.O. Box 581415, Minneapolis, MN 55458-1415 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

(48) Date of publication of this corrected version:
1 April 2004

(15) Information about Correction:
see PCT Gazette No. 14/2004 of 1 April 2004, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMPROVED PSEUDOTYPED RETROVIRUSES

(57) Abstract: Pseudotyped retroviruses having viral glycoproteins with modified O glycosylation regions are provided. Also provided are methods for making the pseudotyped retroviruses of the present invention and for using the pseudotyped retroviruses for transduction of target cells. Cells for stably producing the pseudotyped retroviruses or the present invention are also provided.

WO 2003/102219 A2

IMPROVED PSEUDOTYPED RETROVIRUSES

This application claims the benefit of U.S. Provisional Application Serial No. 60/386,064, filed 04 June 2002, and of U.S. Provisional Application Serial No. 60/458,070, filed 27 March 2003, each of which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

This invention relates generally to pseudotyped viruses and methods of use of the viruses. Specifically, the invention relates to retroviruses pseudotyped with glycoproteins in which an *O*-glycosylation region has been modified, and use of these viruses for gene transfer and gene therapy.

BACKGROUND OF THE INVENTION

Gene therapy is one of the fastest growing areas in experimental medicine. However, most studies are only Phase I or Phase II clinical studies designed mainly to evaluate the toxicity of the viral vectors and constructs being used. A major drawback has been the design of vectors that are both safe and efficacious. Recent efforts in the field have been directed toward the use of retroviral vectors and viral vectors pseudotyped with glycoproteins from highly virulent viruses such as filoviruses.

Retroviruses are ribonucleic acid (RNA) viruses that include an RNA genome enclosed within a viral capsid wherein the capsid is surrounded by an envelope, or lipid bilayer. Glycoproteins present in the lipid bilayer interact with receptors on the surface of various host cells and allow the retroviruses to enter the host cell. Once in the cell, the retroviruses reverse transcribe the RNA of the viral genome into a double-stranded DNA and incorporate the DNA into the cellular genome as a provirus. Gene products from the integrated foreign DNA may then be produced so that progeny viral particles may be assembled. As retroviruses can be modified to carry exogenous nucleotide sequences of interest, such recombinant retroviruses have a variety of uses. For example,

such recombinant retroviruses are important in introducing desired exogenous sequences into a cell, so that relatively high levels of the protein encoded by the sequences may be produced. However, use of such recombinant retroviruses has several drawbacks.

One such drawback is that retroviruses do not have a broad host range. Efforts at increasing the host range of retroviruses have included substituting the envelope glycoproteins of the retrovirus with that of a different virus, thus forming a pseudotyped retrovirus. The pseudotyped retrovirus advantageously has the host range of the different virus. However, some retroviruses have been pseudotyped with viral glycoproteins that are toxic to cells, so the cells can only produce the virus for a limited time. Furthermore, in many cases, the pseudotyped retroviruses cannot be stably produced and may not be produced at a high titer. Stable cell lines have been developed to overcome the toxicity problems and to stably produce such pseudotyped retroviruses. However, there still exists a need for pseudotyped retroviruses that will allow for the production of high titers that would be required for routine gene transfer and/or gene therapy.

Thus it would be desirable to have a pseudotyped retrovirus that is not toxic to cells and produces high titers of a competent virus. It would also be desirable to have a cell line to produce such retroviruses. Methods for using such a virus would also be desirable.

SUMMARY OF THE INVENTION

It has been discovered that deleting the *O*-glycosylation region of a viral glycoprotein of a pseudotyped retrovirus allows for stable production of the pseudotyped virus from various cell lines. Pseudotyped retroviruses containing viral glycoproteins with a deleted *O*-glycosylation region were produced in higher titer than those with wild-type viral glycoproteins, and were also more efficient in transducing target cells. Accordingly, one aspect of the invention provides pseudotyped retroviruses that include recombinant RNA surrounded by, enclosed within, or otherwise associated with a retroviral core that includes a viral capsid, matrix and nucleocapsid. Surrounding the retroviral capsid is a lipid bilayer that includes least one viral glycoprotein disposed therein. The *O*-

glycosylation region of the glycoprotein disposed in the lipid bilayer has been modified through deletion, in whole or in part, or mutation. In one embodiment, the viral glycoprotein is a filoviral glycoprotein such as, but not limited to, Ebola virus or Marburg virus glycoprotein. The recombinant RNA preferably contains a nucleotide sequence that defines a selected biomolecule intended for delivery to the target cell. The nucleotide sequence can encode a protein to be expressed in the target cell, or it can function as a bioactive RNA in the target cell.

In another aspect, the invention provides producer cells for producing pseudotyped retroviruses having a viral glycoprotein in which the *O*-glycosylation region is deleted, in whole or part. Accordingly, the present invention provides eukaryotic producer cells that include nucleotide sequences encoding retroviral proteins Gag polypeptide, Pro polypeptide, and Pol polypeptide and a fourth nucleotide sequence encoding at least one viral glycoprotein in which the *O*-glycosylation region of the glycoprotein has been modified. Preferably, the fourth nucleotide sequence encodes a filoviral glycoprotein, such as, for example, a Marburg virus or Ebola virus glycoprotein having a modified *O*-glycosylation region as described herein. In a preferred form of the invention, the cells stably produce inventive pseudotyped retroviruses.

In a further aspect of the invention, methods for introducing nucleotide sequences into a target cell using the pseudotyped viruses of the present invention are provided. The viruses can be used to introduce a nucleotide sequence into a cell in cell culture, *ex vivo* or *in vivo*. The pseudotyped viruses of the invention are useful for introduction of a nucleotide sequence into a wide range of cell types.

In yet another aspect of the invention methods are provided for producing high titers of pseudotyped retroviruses. The methods involve introducing into a recombinant producer cell nucleotide sequences that encode a viral glycoprotein in which the *O*-glycosylation region is modified. Also present in the cells are nucleotide sequences that encode for other proteins necessary to produce a pseudotyped retrovirus in high titers and that is more efficient in transfecting target cells.

Additional objects, advantages, and features of the present invention will become apparent from the following description, taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is schematic representation of the Ebola virus glycoprotein showing the GP_1 and GP_2 subunits of the glycoprotein drawn to scale with residue numbers indicated below the diagram.

Figure 1B shows the nucleotide sequence (SEQ ID NO:2) and restriction map of plasmid pEboGPΔ309-489. The sequence of the modified Ebola glycoprotein begins with the ATG start codon at nucleotides 956-958 and ends with the TAG stop codon at nucleotides 2447-2449. The *O*-glycosylation region represented by wild-type codons 309-489 has been deleted in this construct. Codon 308 begins at nucleotide 1877 and codon 490 begins at nucleotide 1886. The insertion sequence between codons 308 and 490 (TCTAGA at nucleotides 1880-1886) encodes the dipeptide Ser-Arg.

Figure 1C shows a portion of the nucleotide sequence (SEQ ID NO:1) and restriction map of plasmid pEboGP showing the complete coding sequence of Ebola glycoprotein. The coding sequence begins with the ATG start codon at nucleotides 956-958 and ends with the TAG stop codon at nucleotides 2984-2986. The *O*-glycosylation region represented by wild-type codons 309-489, which is deleted in pEboGPΔ309-489, begins at nucleotide 1880 and ends at nucleotide 2422.

Figure 2 is a western blot showing the expression and incorporation of the Δ308-489 Ebola GP into pseudotyped retrovirus.

Figure 3 is a western blot showing the extent of *O*-glycosylation of the Δ308-489 Ebola GP incorporated into pseudotyped retroviruses.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

It has been discovered that deleting the *O*-glycosylation region of a viral glycoprotein of a pseudotyped retrovirus allows for stable production of the pseudotyped virus from producer cell lines. Pseudotyped viruses having viral glycoproteins with a deleted *O*-glycosylation region were produced in higher

titer than those with wild-type viral glycoproteins and were also more efficient in transducing target cells.

Accordingly, one aspect of the invention provides a pseudotyped retrovirus. The pseudotyped retrovirus is formed from retroviral components that include a retroviral core surrounded by a lipid bilayer, and a recombinant RNA. Disposed in the lipid bilayer is at least one non-wild-type viral glycoprotein which contains a modified *O*-glycosylation region. The modification can take the form of a deletion of the *O*-glycosylation region from the viral glycoprotein, in whole or in part, or mutation of the *O*-glycosylation region. The resulting pseudotyped retrovirus exhibits increased transduction capacity and expanded tropism, compared to the wild-type retrovirus.

In one embodiment, the viral glycoprotein is a filoviral glycoprotein such as, but not limited to, Ebola virus or Marburg virus glycoprotein.

It has been discovered that eukaryotic producer cells may be constructed that stably produce pseudotyped retroviruses having a viral glycoprotein disposed in their lipid bilayer wherein the *O*-glycosylation region of the viral glycoprotein has been modified. Modification of the *O*-glycosylation region results in enhanced glycoprotein processing and incorporation into retroviral particles, allowing for stable production of the pseudotyped viruses.

It has further been discovered that levels of viral transduction into target cells are significantly higher for the pseudotyped retroviruses having glycoproteins with modified *O*-glycosylation regions as compared to pseudotyped retroviruses bearing wild-type glycoproteins.

Glycoprotein for use in retroviral pseudotyping

The present invention contemplates that the *O*-glycosylation region of the glycoprotein used to pseudotype the retrovirus is modified, either by deletion (in whole or part) or mutation. Deletion of the *O*-glycosylation region allows the pseudotyped retrovirus to be stably produced in the recombinant virus producer cells. By "stable production" or "stably produced", it is meant that the producer cells will produce pseudotyped retroviruses indefinitely (i.e., during the life span of the cell). While not wishing to be bound by theory, it is thought that deletion of the *O*-glycosylation region allows for increased

expression of processed glycoprotein in the recombinant virus producer cell. This glycosylation step may be rate limiting during production of the virus, and glycosylation of high amounts of viral glycoproteins may be toxic to the producer cells. Thus, increased expression may be a result of bypassing the post-translation glycosylation step. Increased transduction is observed and may be due to the greater quantity of virus produced.

The viral glycoprotein used in pseudotyping the retrovirus of the invention may be any viral glycoprotein having an *O*-glycosylation region. An *O*-glycosylation region can be identified as one that is rich in proline, serine and threonine residues. Such regions can be discovered by searching protein databases for potential *O*-glycosylation regions using computer algorithms such as NetOGlyc v. 2.0 (J.E. Hansen et al., Glycogonjugate J., 15:115-130, 1998; J.E. Hansen et al., Nucleic Acids Research, 25, 278-282, 1997; J.E. Hansen et al. Biochemical Journal, 308, 801-813, 1995). Alternatively, *O*-glycosylation regions can be identified biochemically, for example by using antibodies to perform western blots.

Filoviral proteins, such as those of Ebola virus (e.g., Ebola Zaire [complete genome, GenBank Acc. No. NC 002549; glycoprotein, GemBank Acc. No. U23187], Ebola Reston [complete genome, GenBank Acc. No. NC 004161; glycoprotein, GenBank Acc. No. U23152] and Ebola Sudan [glycoprotein, GenBank Acc. No. U23069]; see Fig. 1) and Marburg virus [complete genome, GenBank Acc. No. NC 001608; glycoprotein, GenBank Acc. No. Z12132]), have *O*-glycosylation regions and infect a broad spectrum of mammalian hosts. By way of non-limiting example, the *O*-glycosylation region for the Ebola glycoprotein (Zaire) is from about nucleotide 309 to nucleotide 489 of SEQ ID NO:1. The nucleotide sequences encoding the filoviral glycoproteins may be obtained as described in Sanchez et al. (1993) *Virus Res.* 29 (3):215-240 and Will et al., (1993) *J. Virol.* 67:1203-1210. It is expected that other viruses not specifically mentioned herein and having glycoproteins of similar structure to the filoviral glycoproteins may be advantageously used in the present invention.

Viral glycoproteins from filoviruses facilitate the introduction of genes into many different cell types and are therefore preferred for use in the

invention. As an example, Ebola glycoprotein-pseudotyped lentiviruses have specificity in particular for transducing airway epithelia cells from the apical surface and therefore hold promise for reagents for gene therapy for diseases of the lung such as cystic fibrosis.

It will be appreciated by the skilled artisan that the *O*-glycosylation region can be modified by deleting it either in whole or part from the glycoprotein. Deletion involves removing part or all of the nucleotide sequence encoding the region. In embodiments wherein only part of the *O*-glycosylation region is deleted, a sufficient number of nucleotides is deleted to cause the transduction level in a target cell to increase by at least 2-fold, preferably at least 5-fold, and more preferably at least 10-fold, 20-fold, 50-fold or 100-fold. Transduction levels can be readily determined by using the assay described in Example I.

Alternatively, the nucleotide sequence can mutated, for example by modifying it to replace the codons for serine and threonine, the amino acids that are *O*-glycosylated, with codons for conservative amino acid substitutions such as alanine which will not be *O*-glycosylated. A sufficient number of serine and/or threonine residues are mutated such that transduction level into a target cell is increased as described above.

Retroviral particle

The pseudotyped retrovirus of the invention is a recombinant retroviral particle that includes recombinant RNA surrounded by, enclosed within, or otherwise associated with a retroviral core that includes a viral capsid, matrix and nucleocapsid. A viral glycoprotein, modified to delete or mutate the *O*-glycosylation region, is disposed within a lipid bilayer which surrounds the retroviral core.

The recombinant RNA includes both coding and noncoding retroviral control elements. The term "nucleotide sequence", or "nucleic acid sequence" as used herein is intended to refer to a natural or synthetic linear and sequential array of nucleotides and/or nucleosides, and derivatives thereof. The terms "encoding" and "coding" refer to the process by which a nucleotide sequence, through the mechanisms of transcription and translation, provides the

information to a cell from which a series of amino acids can be assembled into a specific amino acid sequence to produce a polypeptide.

Nucleotide sequences required for retroviral transcription, reverse transcription and integration in the target cell are thus included in the recombinant RNA, as well as elements required for incorporation (packaging) of the retrovirus particle (e.g., psi). Reverse transcriptase and integrase give the retrovirus the ability to incorporate a gene encoding a desired protein into a genome of a target cell after the retrovirus contacts, or is incubated with, the cell. The pseudotyped retrovirus may include other proteins, in addition to integrase, that aid its stable integration into the chromosomes of a target cell. For example, with respect to a lentivirus, the pseudotyped retrovirus may include proteins such as vpr, vif and vpu. Other sequences known to the art that are useful for transducing genes may also be present in the recombinant RNA.

An illustrative, nonlimiting example of the structural organization of a recombinant RNA forming a part of the pseudotyped virus of the invention is a nucleotide sequence that begins with an optional 5' long terminal repeat (LTR) or other transcription and/or integration element(s), followed by a primer binding site (PBS) that initiates reverse transcriptase binding, followed by a viral packaging sequence such as psi, which packaging sequence is optionally flanked by splice acceptor and splice donor sites, followed by a reverse transcriptase initiation site for the second strand, followed by an optional 3' LTR. When the recombinant RNA includes a nucleotide sequence that encodes a desired protein or a bioactive RNA (see below), that sequence is typically located between the splice donor site and the reverse transcriptase initiation site.

The pseudotyped retrovirus of the invention can be replicative or nonreplicative. In most applications, it is desirable that the retrovirus be nonreplicative. Replication incompetent retroviruses typically typically lack one or more of the Gag, Pol, Pro or Env proteins.

The recombinant RNA further optionally includes a nucleotide sequence that defines a selected biomolecule intended for delivery to the target cell. The nucleotide sequence can encode a protein to be expressed in the target cell, or it can function as a bioactive RNA in the target cell. This protein or RNA is produced by a target cell upon introduction of the pseudotyped retrovirus into

the target cell. The protein (or RNA) can serve as a detectable marker, or can provide a beneficial or therapeutic effect if introduced into an animal, as described in more detail below.

Recombinant virus producer cells

The invention provides recombinant virus producer cells, preferably eukaryotic cells forming eukaryotic cell lines, that contain nucleotide sequences encoding retroviral Gag polypeptide, retroviral Pro polypeptide, retroviral Pol polypeptide and at least one viral glycoprotein, such as a filoviral glycoprotein, in which the *O*-glycosylation region of the glycoprotein has been deleted or modified, as described above. (Note that *pro* and *pol* often occur so close together on a viral genome that the term *pol* is often used to refer to both *pro* and *pol*). The eukaryotic cell line is grown *in vitro* to produce the pseudotyped retroviruses.

Any of a wide variety of cells can be selected for *in vitro* production of a pseudotyped retrovirus according to the invention. Eukaryotic cells are preferred, particularly mammalian cells including human, simian, canine, feline, equine and rodent cells. Human cells are most preferred. It is further preferred that the cell be able to reproduce indefinitely, and is therefore immortal. Examples of cells that may be advantageously used in the present invention include NIH 3T3 cells, COS cells, Madin-Darby canine kidney cells, human embryonic 293T cells and any cells derived from such cells, such as gpnlslacZ ϕNX cells, which are derived from 293T cells and described in the Examples. Highly transfectable cells, such as human embryonic kidney 293T cells, are preferred. By "highly transfectable" it is meant that at least about 50%, more preferably at least about 70% and most preferably at least about 80% of the cells can express the genes of the introduced DNA.

The retroviral *gag*, *pro* and *pol* nucleotide sequences and other retroviral nucleotide sequences needed for forming the specified pseudotyped retroviruses may be obtained from a wide variety of genera in the family Retroviridae, including, for example, Alpharetrovirus (e.g., avian leukosis virus), Betaretrovirus (e.g., mouse mammary tumor virus), Gammaretrovirus (e.g., murine leukemia virus and Moloney monkey leukemia virus [Mo-MuLV]),

Deltaretrovirus (e.g., bovine leukemia virus), Epsilonretrovirus (walleye dermal sarcoma virus), Lentivirus (e.g., humanimmunodeficiency virus I, feline immunodeficiency virus and Chimpanzee foamy virus) and spumavirus. Retroviral sequences are preferably obtained from the Moloney murine leukemia virus (Mo-MuLV). Such sequences are well known in the art. For example, nucleotide sequences encoding MMLV Gag, Pro and Pol polypeptides may be found in Shinnick et al., *Nature*, 293:543-548 (1981).

Depending on the application, the retroviral components of the pseudotyped retrovirus of the invention are preferably derived from a simple retrovirus such as MoMuLV, or a more complicated retrovirus such as a lentivirus. Examples of lentiviruses include feline immunodeficiency virus (FIV), human immunodeficiency virus (HIV), simian immunodeficiency virus (SIV) and equine infectious anemia virus (EIAV). Unlike most other retroviruses, lentiviruses have the capacity to integrate the genetic material they carry into the chromosomes of non-dividing cells as well as dividing cells. Therefore, lentiviral nucleotide sequences encoding proteins that allow for chromosomal integration of virally transported nucleic acid in non-dividing cells are advantageously employed, as the host range of the pseudotyped retroviruses will be broadened.

Suitable retroviruses, including those described above, are readily publicly available from the American Type Culture Collection (ATCC) and the desired nucleotide sequences may be obtained from these retroviruses by methods known to the skilled artisan. For example, the nucleotide sequences may be obtained by recombinant DNA technology. Briefly, viral DNA libraries may be constructed and the nucleotide sequences may be obtained by standard nucleic acid hybridization or polymerase chain reaction (PCR) procedures, using appropriate probes or primers. Alternatively, supernatant medium from cells infected with the respective virus can be isolated and the desired retroviral nucleotide sequences may be amplified by PCR. Such vectors may also be constructed by other methods known to the art.

It is preferred that the *gag, pro* and *pol* nucleotide sequences are contiguous to each other as found in native retroviral genomes, such as in the order 5'-gag-pro-pol-3'. It is further preferred that these retroviral nucleotide

sequences are chromosomally-integrated into the cellular genome of the producer cell. Furthermore, the gag-pro-pol nucleotide sequences are preferably operably linked at the 5' end of the *gag* nucleotide sequence to a promoter sequence, so that transcription of the sequences may be achieved.

A nucleic acid sequence is "operably linked" to another nucleic acid sequence, such as a promoter sequence, when it is placed in a specific functional relationship with the other nucleic acid sequence. The functional relationship between a promoter and a desired nucleic acid typically involves the nucleic acid and the promoter sequences being contiguous such that transcription of the nucleic acid sequence will be facilitated. Two nucleic acid sequences are further said to be operably linked if the nature of the linkage between the two sequences does not interfere with the ability of the promoter region sequence to direct the transcription of the desired nucleotide sequence, or interfere with the ability of the desired nucleotide sequence to be transcribed by the promoter sequence region. Typically, the promoter element is generally upstream (i.e., at the 5' end) of the nucleic acid coding sequence.

A wide variety of promoters are known in the art, including cell-specific promoters, inducible promoters, and constitutive promoters. The promoters may further be selected such that they require activation by activating elements known in the art, so that production of the protein encoded by the specified nucleic acid sequence may be regulated as desired. It is well within the purview of a person skilled in the art to select and use an appropriate promoter in accordance with the present invention. For example, the promoters that may be advantageously present in the cell, 5' to the gag-pro-pol sequences, include the rat actin promoter and the MMLV promoter. Furthermore, the cytomegalovirus (CMV) promoter has been found to be an excellent promoter in the inventive system.

Other regulatory elements, such as enhancer sequences, cooperate with the promoter and transcriptional start site to achieve transcription of the nucleic acid insert coding sequence, may also be present in the cell 5' to the nucleotide sequences that encode retroviral proteins. By "enhancer" is meant nucleotide sequence elements which can stimulate promoter activity in a cell, such as a bacterial or eukaryotic host cell.

Another aspect of the invention provides methods of making eukaryotic cells for producing pseudotyped retroviruses. The method includes introducing into the cells described above the nucleotide sequences described above, i.e., those encoding the retroviral Gag, Pro and Pol polypeptides, and those encoding a filoviral glycoprotein in which the *O*-glycosylation region is deleted or mutated, in part or in whole, into the cell.

The nucleotide sequences may be introduced into the desired cell utilizing any of variety of vectors known to the skilled artisan. For example, plasmid vectors, cosmid vectors, and other viral vectors, such as retroviral vectors, may be used. It is preferred that the nucleotide sequences encoding the Gag, Pro and Pol polypeptides are on a vector separate from the nucleotide sequences encoding the viral glycoproteins.

In one mode of practicing the invention, plasmid vectors are advantageously used to introduce, or transfect, the nucleotide sequences into the selected cell. A wide variety of plasmid vectors may be used, including pTRE, pCMV-Script and pcDNA3, although pcDNA3 is a preferred vector. The *gag, pro* and *pol* nucleotide sequences are preferably on the same plasmid, and, as discussed above, are preferably contiguous to each other. However, the skilled artisan is aware that other spatial configurations of the nucleotide sequences may be utilized when constructing the plasmids. The vector also preferably includes a promoter 5' to, or upstream from, the *gag* nucleotide sequence. The vectors may further include other regulatory elements, such as enhancer sequences, as discussed above.

The nucleotide sequences encoding the viral glycoproteins are preferably, but need not be, on a plasmid, or other vector, separate from the *gag, pro* and *pol* nucleotide sequences. The viral glycoprotein encoding sequences, such as the sequences encoding the filoviral glycoproteins are also preferably operably linked to a promoter sequence described above. In the case of viral glycoproteins that are formed from separate polypeptides, the two polypeptides may, if desired, be supplied on different plasmids. It is preferred, however, that such sequences are on the same vector, and preferably contiguous to each other so they will be transcribed utilizing the same promoter. In one preferred form of the invention, the promoter sequence is a cytomegalovirus promoter sequence.

Plasmids, or other vectors carrying the nucleotide sequences encoding the viral glycoproteins, may also include other regulatory elements, such as enhancers, as described above.

The vectors may be introduced into the cells in a variety of ways known to the skilled artisan, for example, discussed in *Current Protocols in Molecular Biology,* John Wiley and Sons, edited by Ausubel et al. (1988) and Sambrook et al., *Molecular Cloning, A Laboratory Manual,* Cold Spring Harbor Laboratory Press (2000). For example, vectors may be transfected into a cell by a calcium phosphate precipitation method. Other methods for introduction of the vectors include, for example, electroporation and lipofection.

In order for the producer cells to produce the pseudotyped retroviruses of the invention, they must express the recombinant RNA that is ultimately packaged into the retroviral particles. The producer cells must therefore be supplied with the recombinant RNA. The recombinant RNA may be introduced into the producer cell as DNA or as RNA, in any of a variety of ways known to the skilled artisan. Plasmids are typically used to introduce DNA encoding the recombinant RNA into the producer cells. Alternatively, defective retroviruses (i.e., those which do not have the capability to produce all of the viral proteins necessary for production of a retrovirus having the ability to infect a cell and produce progeny viruses) can be utilized to cause genomic integration of the DNA encoding the recombinant RNA into the producer cell's genome.

Pseudotyped retrovirus as a vehicle for delivery of genetic material

In many cases, one may wish to quickly visually detect those cells which have taken up a vector and that produce a specified protein from the vector. The desired protein encoded by the retroviral recombinant RNA may accordingly be one that facilitates detection of entry of the pseudotyped retrovirus into a target cell. The desired protein may be, for example, a visually detectable component, or marker, such as one that emits visible wavelengths of light, or one that can be reacted with a substrate to produce color of specified wavelengths or other detectable reaction. For example, such nucleotide sequences include the nucleotide sequence encoding the fluorescent proteins such as green, blue and red fluorescent proteins (e.g., *A. victoria* green fluorescent protein (GFP);

nucleotide sequences listed in Prasher et al., *Gene* 111:229 (1992)) and the LacZ gene (produces β-galactosidase, which reacts with a well-known substrate, X-gal), both of which are well known in the art and may be obtained commercially. Cells that have taken up the vector and express the nucleotide sequences encoding a protein may be identified and separated from cells that do not express the sequences by a fluorescence-activated cell sorting procedure as known in the art.

The invention further provides a screening kit for use in evaluating transduction efficiency of a retrovirus in selected target cells. The screening kit includes a pseudotyped retrovirus of the invention, appropriate buffers, and instructions for use.

Advantageously, expression of a detectable marker allows viral entry into a target cell to be confirmed. Expression of the detectable marker can also be used to ensure that the pseudotyped retroviruses that are formed are replication incompetent (i.e., do not have all the sequences necessary in their viral genome to produce progeny retroviruses). Supernatant isolated from target cells transduced by the pseudotyped retroviruses and contacted with a second target test cell should not result in localization of the fluorescent protein in the second target test cell because there should be no retroviruses in the supernatant. A lack of retroviruses in the supernatant would indicate that no replication competent viruses are produced inside the initial target cell, so none of those initial target cells serves as a source of competent retroviruses. Even though it does not make new virus particles, the initial target cell nonetheless expresses the marker because the marker is introduced into its genome via the pseudotyped retrovirus with which it was transduced.

In addition or alternatively to the detectable marker, the pseudotyped retroviruses of the invention may include recombinant RNA that operably encodes a protein that is needed by an animal, either because the protein is no longer produced, is produced in insufficient quantities to be effective in performing its function, or is mutated such that it either no longer functions or is only partially active for its intended function. Other examples of nucleotide sequences that can be included in the retroviral recombinant RNA include those that encode a protein or bioactive RNA (typically a ribozyme) that is effective

to treat a viral infection, stimulate the hosts immune system, or preferentially infect and kill cancer cells in the host.

The present invention further includes methods of introducing desired nucleotide sequences into a target cell using a pseudotyped retrovirus. A target cell is eukaryotic and can be a vertebrate or invertebrate cell. Examples of eukaryotic cells include cells from insects, birds, fish and mammals. Preferably the target cell is a mammalian cell, more preferably a feline, canine, bovine, or human cell. The mammalian cell can be a normal cell or a transformed, cancerous or precancerous cell. The target cell may be a cell already infected by a virus.

In one embodiment, the method of delivering the nucleotide sequences to the target cell includes contacting, or transducing, a target cell permissive for filoviral entry, with a retrovirus that has been pseudotyped with a filoviral glycoprotein as described above that includes the desired nucleotide sequence in its genome. The cells can be contacted while present in the organism (*in vivo*) (Sharkey et al., *J Virol*, 75:2653-2659 (2001)), while explanted and outside the organism (*ex vivo*), or in cell culture (Sinn et al., *J Virol* 2003 May 15;77(10):5902-10 and Kang et al., *J. Virol.*, 76:9378-9388 (2002)). Cells contacted *ex vivo* can, if desired, be reimplanted into the organism. *In vivo* delivery of the pseudotyped retrovirus can be accomplished using methods well known to the art of therapeutic delivery of genetic material, such as injection, perfusion, aerosolization followed by inhalation, and the like. Dosages can be readily determined by reference to existing retroviral gene therapies.

When the nucleotide sequence encodes a desired protein, the target cell is selected so that it allows expression of the selected nucleotide sequence. The level of transduction may be determined by assay methods known to the skilled artisan, and include assaying for the protein of interest encoded by the introduced nucleotide sequences or assaying for the presence of the nucleotide sequences.

Any existing retroviral therapy can be modified according to the invention by including in the lipid bilayer of the retrovirus the modified glycoproteins described herein. Pseudotyping with the modified glycoproteins

increases the efficiency of gene therapy because the transduction levels are significantly increased, making the procedure more economical as well.

Reference will now be made to specific examples illustrating the compositions and methods above. It is to be understood that the examples are provided to illustrate preferred embodiments and that no limitation to the scope of the invention in intended thereby.

Example I. Murine Leukemia Virus Pseudotyped with Ebola Glycoprotein Lacking *O*-Glycosylation Region

MATERIALS AND METHODS

Cell lines and culture conditions. The human kidney cell line 293 (ATCC Number CRL-1573), the mouse embryo cell line NIH 3T3 (CRL-1658), and the 293T-derived ϕNX (second generation retroviral packaging cells) (Grignani et al., *Cancer Res*, 58:14-19 (1998); Pear et al., *Proc. Natl. Acad. Sci. USA*, 90:8392-8396 (1993); Swift et al., p. 10.17.14-10.17.29. In R. Coico (ed.), Current Protocols in Immunology supp. 31. J. Wiley & Sons (1999)) and gpnlslacZ cell lines were cultured in Dulbecco's minimal essential medium (DMEM) containing 10% heat inactivated fetal bovine serum, 2 mM glutamine, 100 units Penicillin G, and 100 µg/ml streptomycin sulfate, with or without 0.25 µg/ml amphotericin B (growth medium). The gpnlslacZ cells produce envelope protein-deficient replication-incompetent Mo-MuLV particles carrying MFG.S-nlslacZ, a retroviral vector encoding a nuclear localizing β-galactosidase (Sharkey et al., *J Virol*, 75:2653-2659 (2001)).

Plasmids and site-directed mutagenesis. A modified version of the plasmid pTM1 was used in transient expression studies of GP sequences using a vaccinia virus-T7 RNA polymerase (VV-T7) system (Elroy-Stein et al., *Proc Natl Acad Sci U S A*, 86:6126-6130 (1989)). The pTM1 vector was modified to remove an ATG codon (within an *Nco*I site) at the beginning of the multiple cloning site by *Nco*I digestion, mungbean nuclease treatment, and ligation of the blunt-ended DNA. This vector, pTM1(Δ*Nco*I), was used to subclone the entire Ebola virus GP open reading frame (ORF). The GP ORF was cleaved from the

buffer) for 30 minutes at room temperature. Membranes were washed as before, then treated with a commercial chemiluminescent substrate solution (Amersham Pharmacia Biotech), according to the protocols provided by the manufacturer. Specific reactivity to GP was visualized by exposing treated blots to X-ray film.

RESULTS

Pseudotyped retroviruses bearing GPs with altered glycosylation. The role of *O*-glycosylation of the Ebola virus GP was examined through analysis of the effects of deleting the region of the protein that is *O*-glycosylated (Jeffers et al., *J Virol* 2002 Dec;76(24):12463-72). Remarkably, processing and viral incorporation of the Δ309-489 GP was greatly enhanced as shown in Figure 2. The migration of the mature GP_1, GP_0 (the glycosylated but uncleaved glycoform), GP_{pre} (the *N*-glycosylated but not *O*-glycosylated uncleaved glycoform) and deglycosylated GP_0 and GP_{pre} forms of the wild-type GP and of the GP_1, GP_0 and deglycosylated GP_1 and GP_0 forms of the Δ309-489 GP is indicated. There was also a corresponding increase of 696 ± 142% in transduction by the Δ309-489 GP pseudotyped viruses as compared to wild type. The absence of an increase in the mobility of the Δ309-489 GP upon sialadase A and endo-*O*-glycosidase treatment provides confirmation that the region of *O*-glycosylation of the glycoprotein has been removed (Figure 3). The migration of the mature GP_1 of the wild-type and Δ309-489 GPs is indicated. In this experiment a glycosylated serum protein possessing a mobility intermediate between those of the wild-type and Δ309-489 GP_1s was detected. The heterogeneous mobility of the PNGaseF-treated proteins is indicative of incomplete removal of *N*-glycosylation.

The effect of deleting the *O*-glycosylation region of GP_1 (Δ309-489) on expression and transduction were striking. This segment, which is rich in proline, serine, and threonine residues is the most variable among the Ebola GPs. Elimination of this mucin-like region results in enhanced GP processing and incorporation into retroviral particles (Figure 2) and consequently higher levels of transduction by the pseudotyped retroviruses. It is possible that the wild-type GP is retained in the Golgi apparatus until all of the serines and

plasmid pGEM-EMGP1 (Sanchez et al., *Virus Res*, 29:215-240 (1993)) by digestion with *Bam*HI and *Dra*I, and the fragment isolated and directionally ligated into the pTM1(Δ*Nco*I) vector cleaved with *Bam*HI and *Stu*I. The resulting clone, pTM1(Δ*Nco*I)-GP, was used as the target DNA for all site-directed mutagenesis reactions. This clone encodes a GP sequence that differs from the wild-type amino acid sequence in a single residue within the membrane spanning sequence (I662V), and for comparative purposes will be referred to as "wild-type sequence". This mutation is present in the original pGEM3Zf(-)-GP clone, but does not appear to affect the processing or function of the GP. GP residue numbering commences with the methionine of the signal sequence and is continuous through the GP_1 and GP_2 sequences.

The GP clone in which the mucin region was deleted (Δ309-489) was generated from two PCR clones linked by an *Xba*I restriction site, which resulted in the replacement of the mucin sequence with two residues (serine-arginine). Mutations in isolated plasmid clones were identified by direct sequencing of mini-prep DNA using dye-terminator cycle sequencing reactions (ABI) analyzed on either an ABI 373 or 377 sequencer. Large-scale preparations for each type of mutated plasmid DNA were made using commercial kits (Promega Corp. or 5 Prime → 3 Prime, Inc.). The DNA was quantified by UV_{260} absorbance readings, and then stored at -70°C until needed. The coding region (*Bam*HI/*Sal*I fragments) from the plasmid pTM1(ΔNcoI)-GP and mutated versions of this DNA were separately ligated into the *Bam*HI/*Xho*I polylinker sites of the vector pcDNA3 (Invitrogen) to yield pEboGP (Fig. 1C) and pEboGPΔ309-489 (Fig. 1B). These plasmids were then cloned in *E. coli*, and plasmid DNA was isolated for use in pseudotyping studies.

Retrovirus pseudotyping and virus transduction assays. Pseudotyped retrovirus particles consisting of the MuLV cores and the Ebola GP in their envelopes were produced by transfecting wild-type or mutated plasmid DNA into gpnlslacZ cells as previously described by Sharkey et al. (*J. Virol.* 75:2653-2659 (2001)). Virus transduction of β-galactosidase activity into NIH 3T3 cells was determined as described by Sharkey et al. (*J. Virol.* 75:2653-2659 (2001)). All data presented are the average of the results of at least three experiments.

Immunoblot analysis of Ebola virus glycoprotein expression, processing, and incorporation into pseudotyped retroviruses. Medium from transfected ϕNX cells (Grignani et al., *Cancer Res* 58:14-19 (1998); Sanchez et al., *J Virol* 72:6442-6447 (1998); Swift et al., p. 10.17.14-10.17.29. In R. Coico (ed.), Current Protocols in Immunology supp. 31. J. Wiley & Sons (1999)) containing recombinant retroviruses were passed through a 0.45 μm filter and centrifuged through a 30% sucrose cushion at 25,000 RPM in a Beckman 50.2-Ti rotor in a Beckman SS-71 centrifuge. The fluid was aspirated from centrifuge tubes and discarded, and the virus pellet was suspended in 100μl of RIPA buffer (140 mM NaCl, 10 mM Tris HCl pH 8.0, 5 mM EDTA, 1% Na deoxycholate, 1% Triton X-100, 0.1% SDS). Cells were treated with lysis buffer (50 mM Tris HCl pH 8.0, 5 mM EDTA, 150 mM NaCl, 1% Triton X-100), and the cell lysates were then centrifuged in a microcentrifuge at 16,100 x *g* for ten minutes. The proteins in the cell lysate and the suspended viral pellet were each precipitated with a final concentration of 4% TCA for two minutes. The precipitated proteins were then centrifuged at 16,100 x *g* in a microcentrifuge for ten minutes. The supernatant fluid was aspirated and discarded. The pellet was suspended in an equal volume of 1 M Tris[hydroxymethyl]-aminomethane and vortexed vigorously. Proteins whose glycosylation was analyzed were treated sequentially with PNGase F, which removes *N*-linked glycosylation, and with both Sialidase A and Endo-*O*-glycosidase (ProZyme, Inc.), which together remove *O*-linked glycosylation, following protocols provided by the supplier. The pellet suspension was then mixed with 1/6 the volume of 300 mM Tris pH 6.8, 60% glycerol (w/v), 4% SDS (w/v), 0.0012% bromophenol blue (w/v), 6% 2-mercaptoethanol (v/v) and boiled for 5 minutes.

Equal amounts of proteins as determined by the Bradford assay were separated by SDS-PAGE (8.5% acrylamide), and electrophoretically blotted onto nitrocellulose membranes. Membrane blots were immersed in reaction buffer (20 mM Tris, 137 mM NaCl, 0.1% Tween 20, pH 7.6) containing 1% bovine serum albumin and incubated overnight at 4°C. Blots were incubated in reaction buffer containing a rabbit-anti-Ebola SGP/GP diluted 1:1000 for 1 hour at room temperature, washed 3 times in reaction buffer, and reacted with a Goat anti-rabbit-horseradish peroxidase conjugate (diluted 1:20,000 in reaction

threonine residues in the mucin region are modified. While not wishing to be bound by theory, it is thought that elimination of this segment may permit more rapid transit through the Golgi apparatus and higher levels of processing to $GP_1$and GP_2 and of cell-surface expression. Increased viral incorporation may also result from a diminution of GP toxicity. It has been reported that the deletion of the *O*-glycosylation region reduces the cytopathic effects of Ebola virus GP expression (Yang et al., *Science*, 279:1034-1037 (1998)). It has also been suggested that the expression of high levels of the wild-type Ebola GP might lead to exhaustion of the cellular glycosylation machinery (Volchkov et al., *Science*, 291:1965-9 (2001)), which is consistent with the present results and present interpretation.

Example II. Expression of Green Fluorescent Protein in Target Cells Transduced with Pseudotyped Retrovirus

A plasmid was constructed based on pcDNA3.1 (Invitrogen) expresses the Ebola virus (Zaire strain) glycoprotein with its *O*-glycosylation region deleted (amino acids 309-489), under control of the cytomegalovirus (CMV) promoter. pcDNA3.1 is similar to pcDNA1 but it also contains zeocin resistance. This plasmid was transfected into human cell lines (gpGFP, Taylor et al., *Mol. Biol. Cell*, 10:2803-2815 (1999)) that have also been transfected with genes encoding the Moloney murine leukemia virus (Mo-MuLV) *gag* and *pol* (including *pro*) genes, the plasmid MFG.S-GFP, and a gene encoding the *A. victoria* green fluorescent protein (GFP).

Recombinant pseudotyped retrovirus recovered from the supernatant medium of such cells was incubated with several different cell lines (a murine cell line, NIH3T3; a human cell line, HeLa; and a hamster cell line, BHK) and was shown to be capable of introducing the gene encoding the green fluorescent protein into the target cells. Transduction levels associated with the modified pseudovirus were 4.7 fold higher than with the retroviruses pseudotyped with unmodified (native) Ebola virus glycoprotein. These recombinant modified Ebola glycoprotein-pseudotyped retroviruses have substantially improved titers

that make *in vivo* gene transfer and gene therapy experiments with such viruses feasible for the first time.

Example III. Feline Immunodeficiency Virus Pseudotyped with Ebola Glycoprotein Lacking *O*-Glycosylation Region

Feline immunodeficiency virus (FIV)-based vectors, which are a non-primate lentivirus vectors, were pseudotyped using envelope glycoproteins (GPs) from the filoviruses Marburg and Ebola virus (Sinn et al., J Virol 2003 May 15;77(10):5902-10). We observed that primary cultures of well-differentiated human airway epithelia were transduced when filovirus GP pseudotyped FIV was applied to the apical surface. Furthermore, by deleting a heavily *O*-glycosylated extracellular region of the Ebola GP, we improved the titer of concentrated vector several fold.

Vector production. The second-generation FIV vector system utilized in this study was reported previously (Johnston et al., *J. Virol.* 73:4991-5000 (1999) and Wang et al., *J. Clin. Inv.* 104:R55-62 (1999)). The FIV vector construct expressed the β-galactosidase cDNA directed by the CMV promoter. EBOΔO (pEZGP 309-489) has been previously described (Jeffers et al., *J. Virol.*, 76:12463-12472 (2002); Example I). Pseudotyped FIV vector particles were generated by transient transfection of plasmid DNA into 293T cells as described previously (Johnston et al., *J. Virol.* 73:4991-5000 (1999)). FIV vector preparations were titered on HT1080 cells at limiting dilutions and these titers were used to calculate the multiplicities of infection (MOIs). In addition, we found that the filoviral glycoprotein conferred enough stability to the lentiviral vector to withstand centrifuge concentration of greater 1000-fold (data not shown); however, we typically concentrated vector 250-fold by centrifugation for *in vitro* experiments.

Deletion of the O-glycosylated region from the extracellular domain of filoviral glycoproteins. An initial strategy for enhancing filoviral pseudotyped FIV-vector titer was to delete an expansive region from the extracellular domain thought to be heavily *O*-glycosylated. By deleting this region, the efficiency of envelope protein synthesis and transport to the cell surface is enhanced

(Example I). This region may be functionally less important than the flanking regions of the protein simply because of little sequence conservation in this region among all filoviral isolates. The deletion of amino acids 309-489 from the Ebola glycoprotein (EBOΔO) resulted in a marked 74-fold increase in titer over the average titer obtained with the wild-type Ebola glycoprotein.

The complete disclosure of all patents, patent applications, and publications, and electronically available material (including, for example, nucleotide sequence submissions in, e.g., GenBank and RefSeq, and amino acid sequence submissions in, e.g., SwissProt, PIR, PRF, PDB, and translations from annotated coding regions in GenBank and RefSeq) cited herein are incorporated by reference. The foregoing detailed description and examples have been given for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The invention is not limited to the exact details shown and described, for variations obvious to one skilled in the art will be included within the invention defined by the claims.

What is claimed is:

1. A pseudotyped retrovirus comprising recombinant RNA associated with a retroviral core surrounded by a lipid bilayer having disposed therein a glycoprotein comprising a modified *O*-glycosylation region, the recombinant RNA comprising (i) a nucleotide sequence defining a selected biomolecule intended for delivery to a target cell, and (ii) retroviral control elements for packaging, reverse transcription and integration of the retrovirus into a target cell.

2. The pseudotyped retrovirus of claim 1 wherein the retroviral core and control elements are from Moloney murine leukemia virus (Mo-MuLV).

3. The pseudotyped retrovirus of claim 3 wherein the retroviral core and control elements are from a lentivirus.

4. The pseudotyped retrovirus of claim 1 wherein the lentivirus is feline immunodeficiency virus (FIV), human immunodeficiency virus (HIV), simian immunodeficiency virus (SIV) or equine infectious anemia virus (EIAV).

5. The pseudotyped retrovirus of claim 1 wherein the glycoprotein is a filovirus glycoprotein.

6. The pseudotyped retrovirus of claim 1 wherein the selected biomolecule is a protein.

7. The pseudotyped retrovirus of claim 1 wherein the selected biomolecule is a bioactive RNA.

8. The pseudotyped retrovirus of claim 1 having a transduction efficiency into target cells of at least 2-fold higher than a retrovirus pseudotyped with the wild-type glycoprotein.

9. A pseudotyped retrovirus comprising recombinant RNA associated with a retroviral core surrounded by a lipid bilayer having disposed therein an Ebola glycoprotein comprising a modified *O*-glycosylation region, the recombinant RNA comprising (i) a nucleotide sequence defining a selected biomolecule intended for delivery to a target cell, and (ii) retroviral control elements for packaging, reverse transcription and integration of the retrovirus into a target cell.

10. The pseudotyped retrovirus of claim 9 wherein the Ebola glycoprotein contains a deletion of nucleotides 309 to 489 in SEQ ID NO:1.

11. The pseudotyped retrovirus of claim 10 wherein the retroviral core and control elements are from Mo-MuLV retrovirus.

12. The pseudotyped retrovirus of claim 10 wherein the retroviral core and control elements are from a lentivirus.

13. A pseudotyped retrovirus pseudotyped with a glycoprotein comprising a modified *O*-glycosylation region, the pseudotyped retrovirus having a transduction efficiency into a target cell of at least 2-fold higher than a retrovirus pseudotyped with the wild-type glycoprotein.

14. A recombinant virus producer cell comprising *gag*, *pro* and *pol* nucleotide sequences and a nucleotide sequence encoding a glycoprotein comprising a modified *O*-glycosylation region.

15. The recombinant virus producer cell of claim 14 wherein the glycoprotein is an Ebola glycoprotein containing a deletion of nucleotides 309 to 489 in SEQ ID NO:1.

16. The recombinant virus producer cell of claim 15 which is a NIH 3T3 cell, COS cell, Madin-Darby canine kidney cell, human embryonic 293T cell or any cell derived therefrom.

17. A method for making a pseudotyped retrovirus comprising supplying a recombinant RNA to the recombinant virus producer cell of claim 12, wherein recombinant RNA comprises (i) a nucleotide sequence defining a selected biomolecule intended for delivery to a target cell, and (ii) retroviral control elements for packaging, reverse transcription and integration of the retrovirus into a target cell, under conditions such that pseudotyped retrovirus is produced.

18. The method of claim 17 wherein supplying the recombinant RNA to the producer cell comprises introducing a DNA encoding the recombinant RNA into the producer cell.

19. The method of claim 17 supplying the recombinant RNA to the producer cell comprises introducing the recombinant RNA into the cell.

20. A method for transducing a target cell comprising contacting a target cell with the pseudotyped retrovirus of claim 1.

21. The method of claim 18 wherein the target cell is an insect cell, a bird cell, a fish cell or a mammalian cell.

22. The method of claim 19 wherein the target cell is a human cell.

23. The method of claim 18 wherein the cell is *in vivo*, *ex vivo*, or in cell culture.

24. The method of claim 18 wherein the selected biomolecule is a protein, and wherein the transduced target cell expresses the protein.

25. The method of claim 18 wherein the selected biomolecule is a bioactive RNA, and wherein the transduced target cell produces the bioactive RNA.

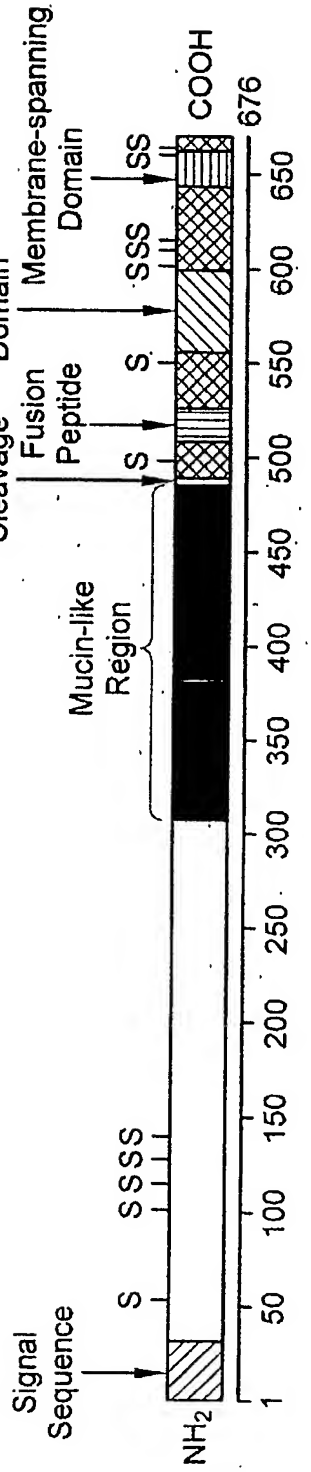
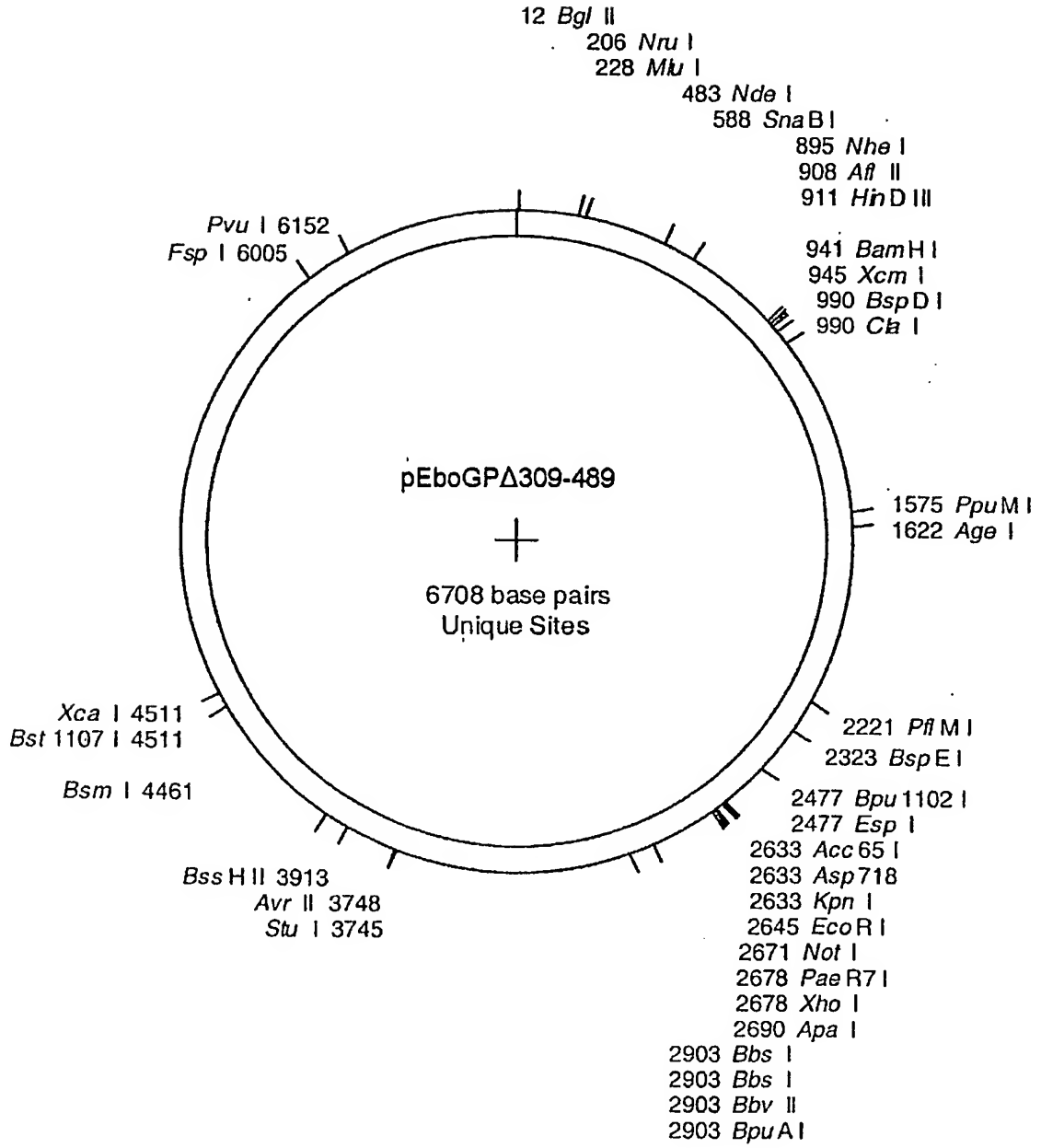
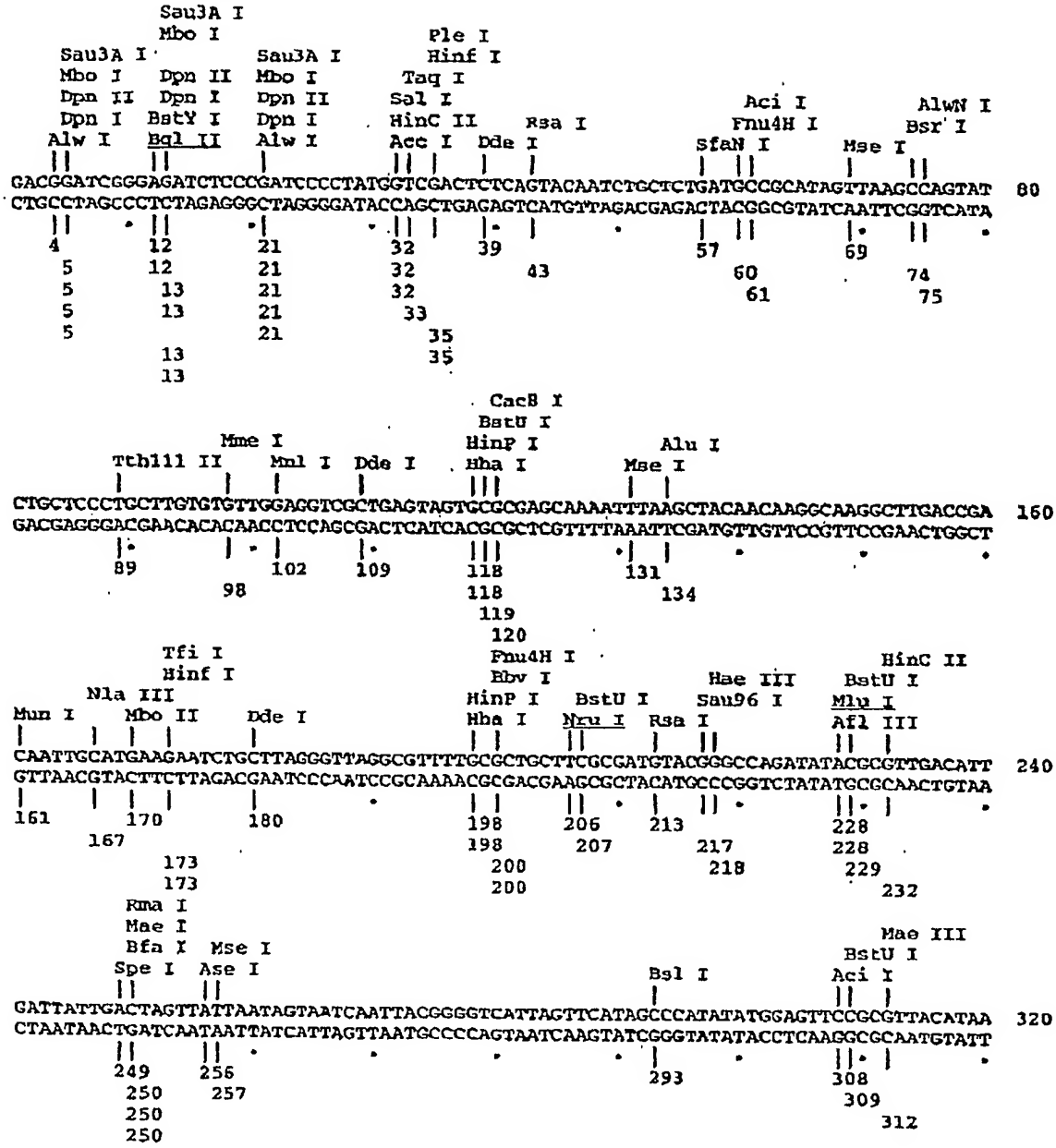
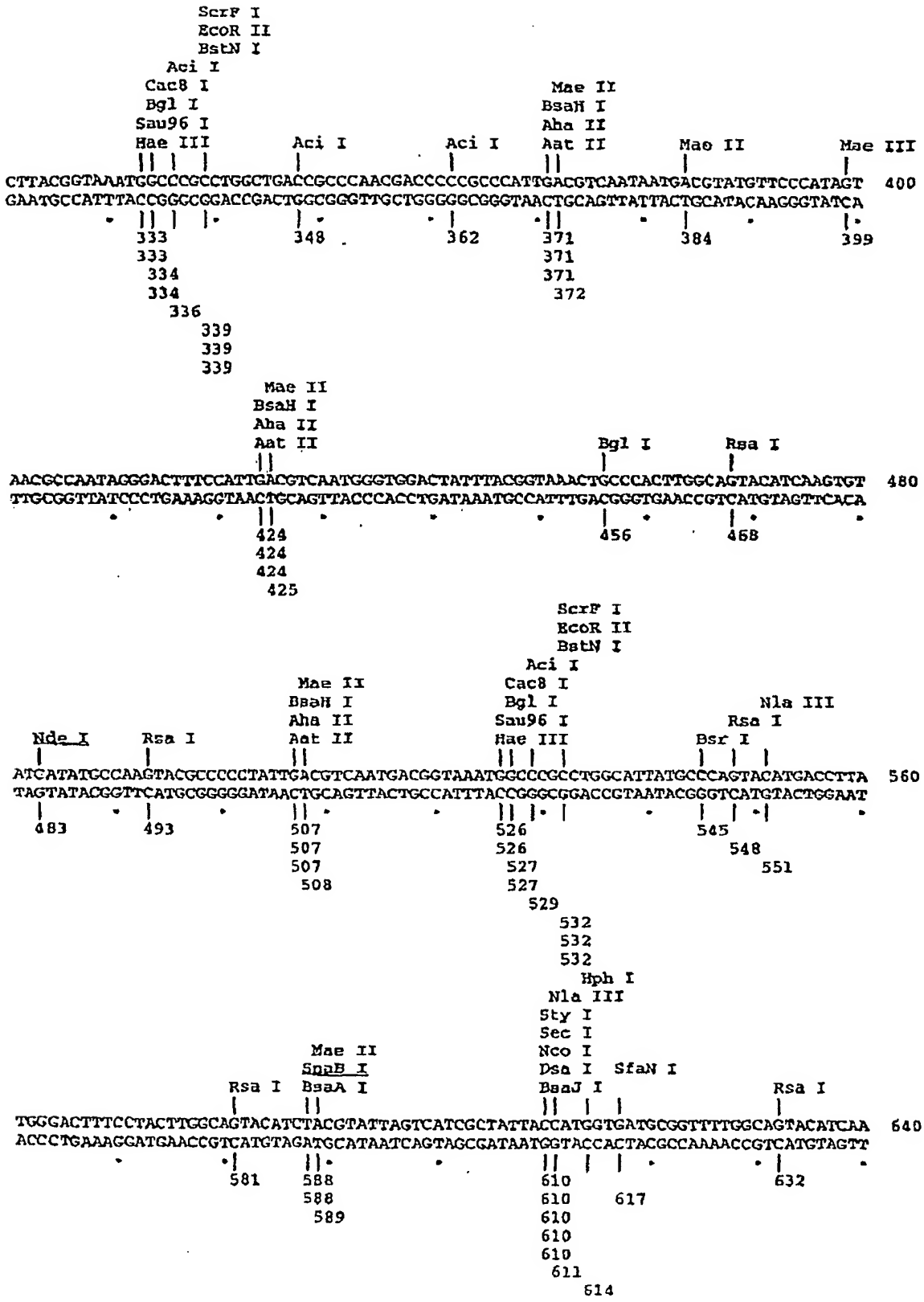

GP1
GP2
Signal Sequence
Furin Cleavage
Coiled-coil Domain
Fusion Peptide
Membrane-spanning Domain
Mucin-like Region
S
SSSS
SSS
SS
S
NH2
COOH
1
50
100
150
200
250
300
350
400
450
500
550
600
650
676

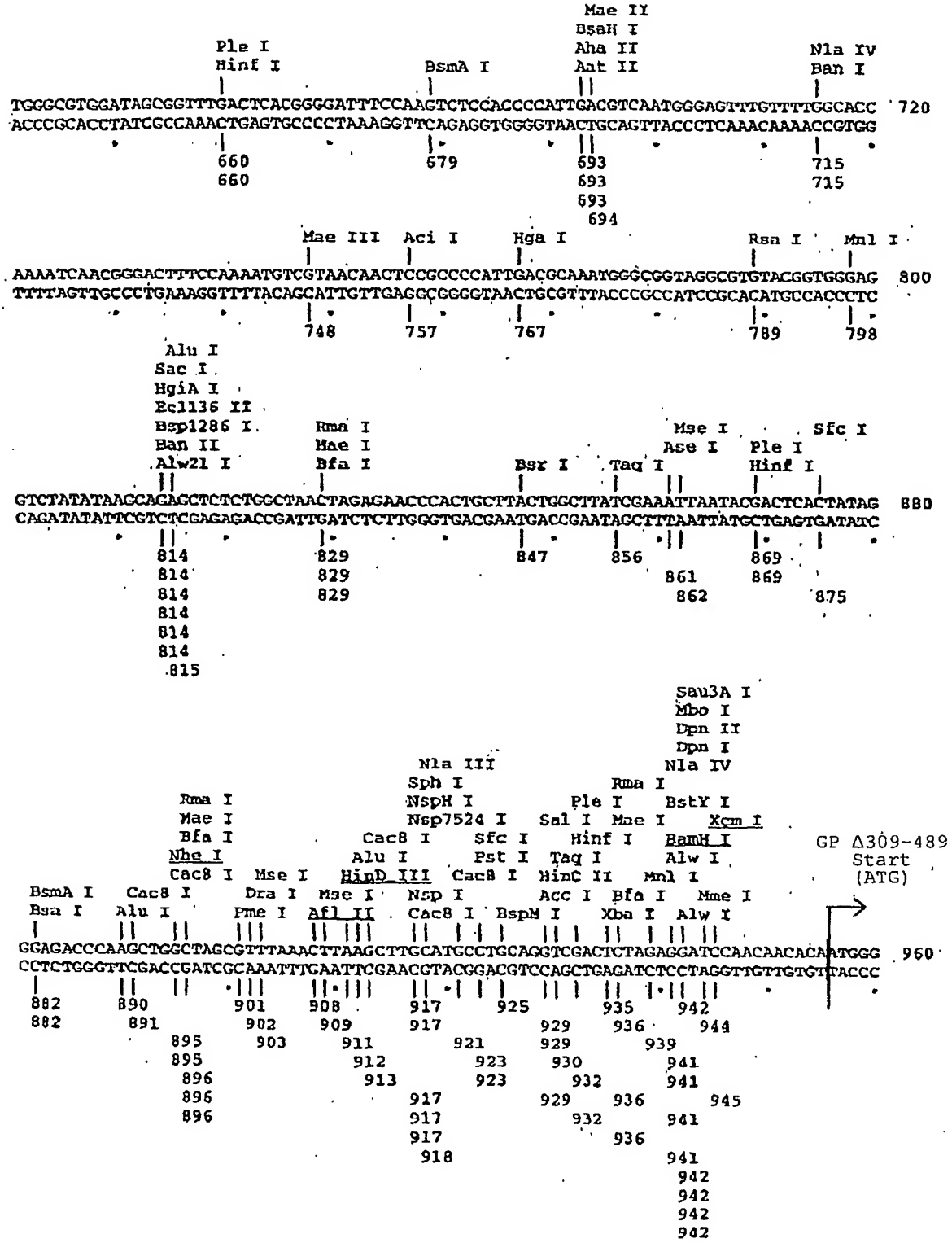
Fig. 1A

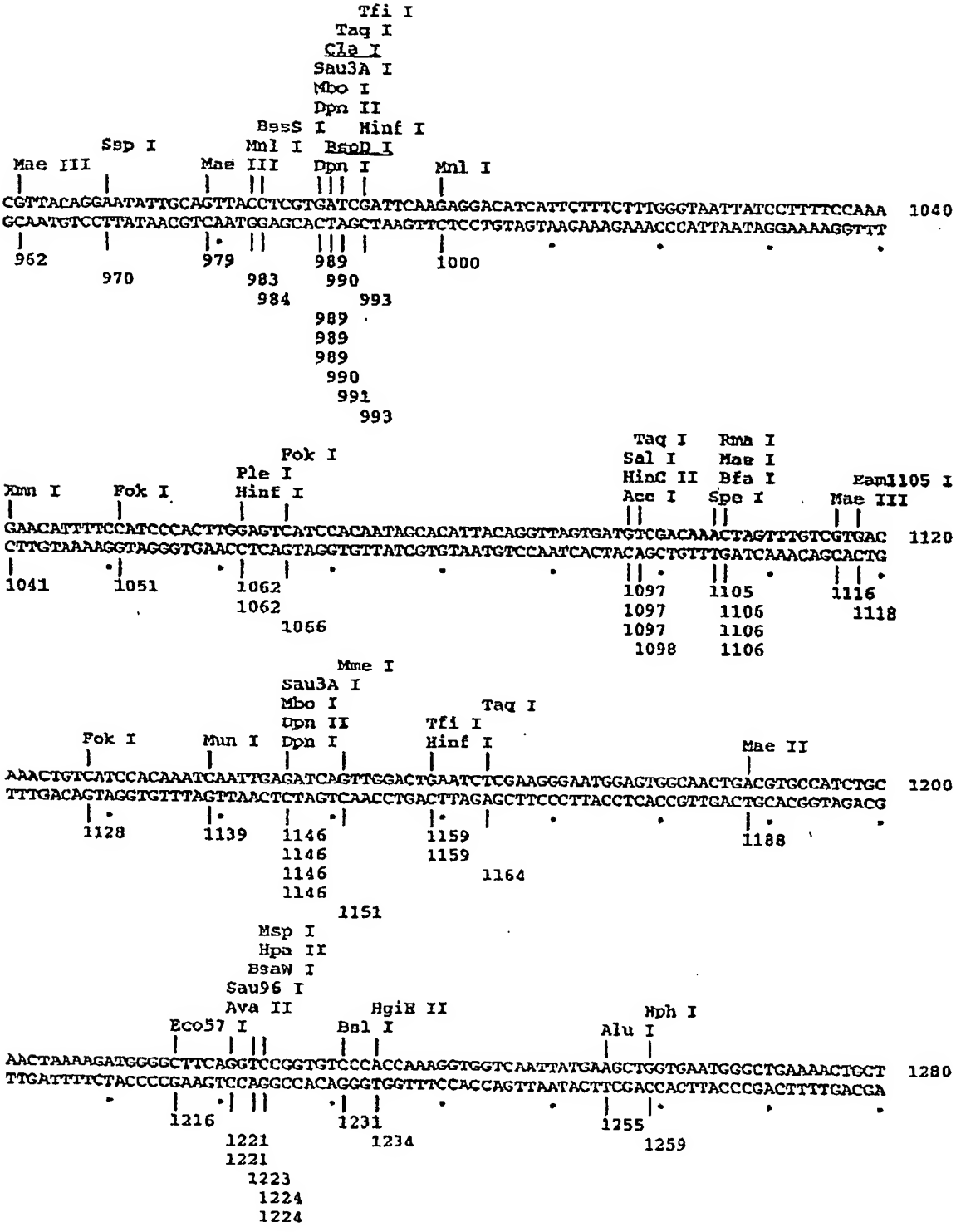
Fig. 1B

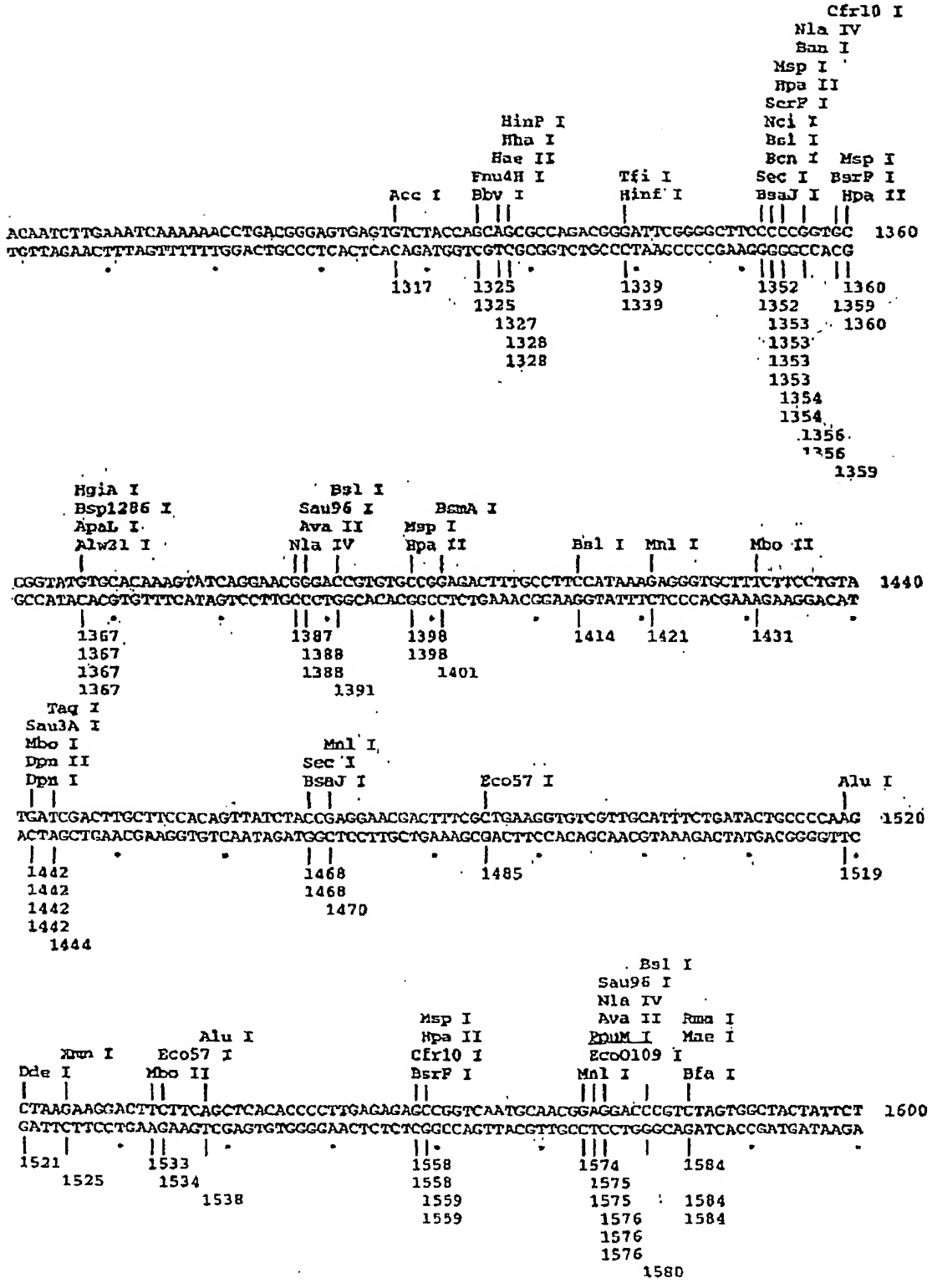

12 Bgl II
206 Nru I
228 Mlu I
483 Nde I
588 SnaB I
895 Nhe I
908 Afl II
911 HinD III
941 BamH I
945 Xcm I
990 BspD I
990 Cla I
1575 PpuM I
1622 Age I
2221 Pfl M I
2323 BspE I
2477 Bpu1102 I
2477 Esp I
2633 Acc65 I
2633 Asp718
2633 Kpn I
2645 EcoR I
2671 Not I
2678 PaeR7 I
2678 Xho I
2690 Apa I
2903 Bbs I
2903 Bbs I
2903 Bbv II
2903 BpuA I
Bss H II 3913
Avr II 3748
Stu I 3745
Xca I 4511
Bst 1107 I 4511
Bsm I 4461
Pvu I 6152
Fsp I 6005
pEboGPΔ309-489
6708 base pairs
Unique Sites

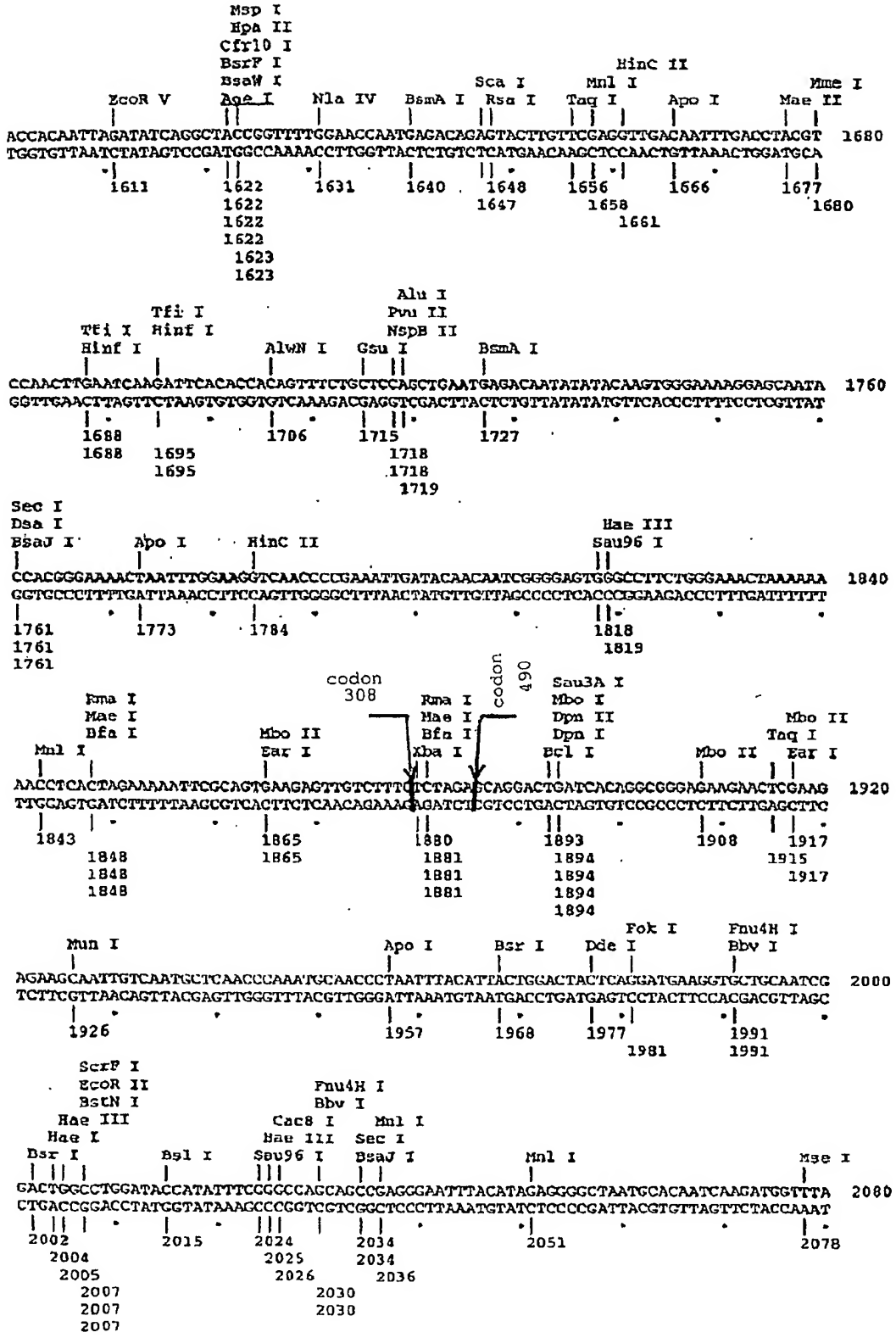

Sau3A I
Mbo I
Ple I
Hinf I
Sau3A I
Mbo I
Dpn II
Dpn I
BstY I
Bgl II
Sau3A I
Mbo I
Dpn II
Dpn I
Alw I
Taq I
Sal I
HinC II
Acc I
Rsa I
Dde I
Aci I
Fnu4H I
SfaN I
AlwN I
Bsr I
Mse I
GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTAT 80
CTGCCTAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATA
4 12 21 32 39 57 69
5 12 21 32 43 60 74
5 13 21 32 61 75
5 13 21 33
5 21 35
13 35
13
CacS I
BstU I
Mme I
HinP I
Alu I
Tth111 II
Mnl I
Dde I
Hha I
Mse I
CTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGA 160
GACGAGGGACGAACACACAACCTCCAGCGACTCATCACGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAACTGGCT
89 102 109 118 131
98 118 134
119
120
Tfi I
Hinf I
Fnu4H I
Bbv I
HinC II
BstU I
Nla III
HinP I
BstU I
Hae III
Sau96 I
Mlu I
Mun I
Mbo II
Dde I
Hha I
Nru I
Rsa I
Afl III
CAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATT 240
GTTAACGTACTTCTTAGACGAATCCCAATCCGCAAAACGCGACGAAGCGCTACATGCCCGGTCTATATGCGCAACTGTAA
161 170 180 198 206 213 228
167 198 207 217 228
173 200 218 229
173 200 232
Rma I
Mae I
Mae III
Bfa I
Mse I
BstU I
Spe I
Ase I
Bsl I
Aci I
GATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAA 320
CTAATAACTGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATT
249 256 293 308
250 257 309
250 312
250

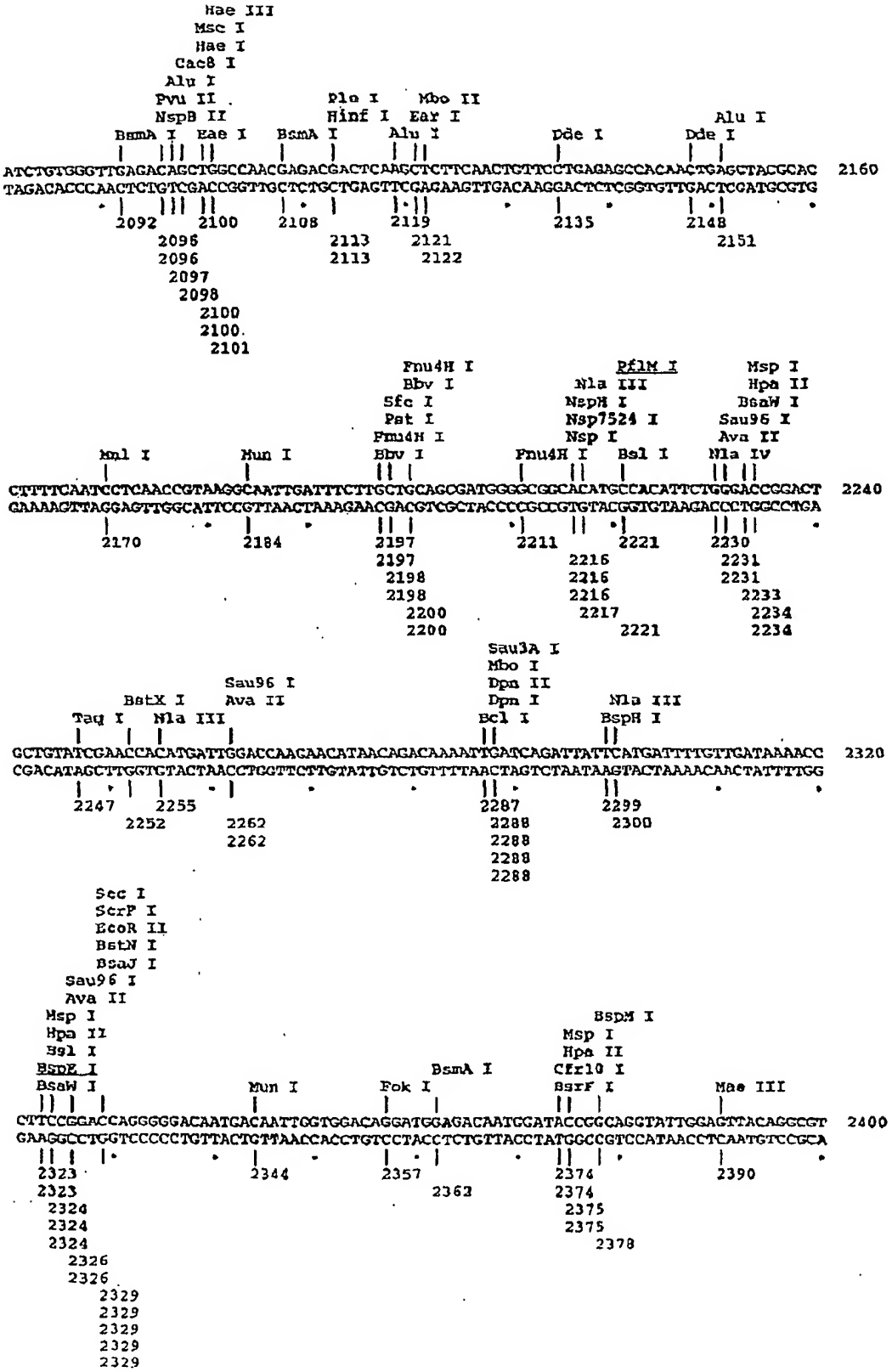

ScrF I
EcoR II
BstN I
Aci I
Cac8 I
Bgl I
Sau96 I
Hae III
Aci I
Aci I
Mae II
BsaH I
Aha II
Aat II
Mae II
Mae III
CTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT 400
GAATGCCATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCATACAAGGGTATCA
333
333
334
334
336
339
339
339
348
362
371
371
371
372
384
399
Mae II
BsaH I
Aha II
Aat II
Bgl I
Rsa I
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGT 480
TTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACA
424
424
424
425
456
468
ScrF I
EcoR II
BstN I
Aci I
Cac8 I
Bgl I
Sau96 I
Hae III
Mae II
BsaH I
Aha II
Aat II
Nla III
Rsa I
Bsr I
Nde I
Rsa I
ATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTA 560
TAGTATACGGTTCATGCGGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAAT
483
493
507
507
507
508
526
526
527
527
529
532
532
532
545
548
551
Hph I
Nla III
Sty I
Sec I
Nco I
Dsa I
SfaN I
Mae II
SnpB I
BsaA I
BsaJ I
Rsa I
Rsa I
TGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAA 640
ACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTT
581
588
588
589
610
610
610
610
610
611
614
617
632

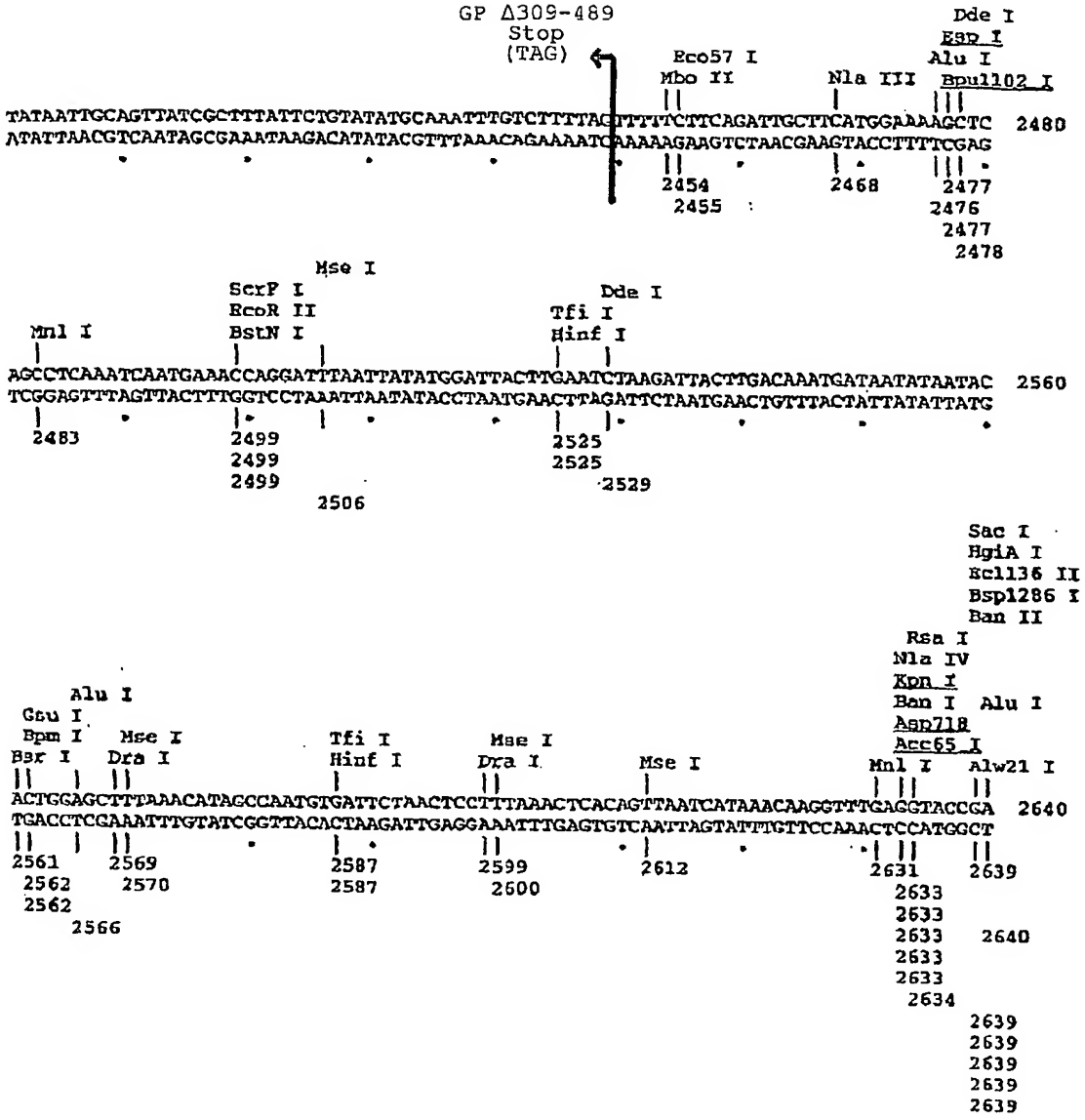

Mae II
BsaH I
Aha II
Ant II
Ple I
Hinf I
BsmA I
Nla IV
Ban I
TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC 720
ACCCGCACCTATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGG
660 679 693 715
660 693 715
693
694
Mae III
Aci I
Hga I
Rsa I
Mnl I
AAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAG 800
TTTTAGTTGCCCTGAAAGGTTTTACAGCATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTC
748 757 767 789 798
Alu I
Sac I
HgiA I
Ecl136 II
Bsp1286 I
Ban II
Alw21 I
Rma I
Mae I
Bfa I
Bsr I
Taq I
Mse I
Ase I
Sfc I
Ple I
Hinf I
GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAG 880
CAGATATATTCGTCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATC
814 829 847 856 869
814 829 861 869
814 829 862 875
814
814
814
815
Sau3A I
Mbo I
Dpn II
Dpn I
Nla III
Nla IV
Sph I
Rma I
NspH I
Ple I
BstY I
Rma I
Mae I
Nsp7524 I
Sal I
Mae I
Xcm I
Bfa I
Cac8 I
Sfc I
Hinf I
BamH I
GP Δ309-489
Nhe I
Alu I
Pst I
Taq I
Alw I
Start
Cac8 I
Mse I
HinD III
Cac8 I
HinC II
Mnl I
(ATG)
BsmA I
Cac8 I
Dra I
Mse I
Nsp I
Acc I
Bfa I
Mme I
Bsa I
Alu I
Pme I
Afl II
Cac8 I
BspM I
Xba I
Alw I
GGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCAACAACACAATGGG 960
CCTCTGGGTTCGACCGATCGCAAATTTGAATTCGAACGTACGGACGTCCAGCTGAGATCTCCTAGGTTGTTGTGTTACCC
882 890 901 908 917 925 935 942
882 891 902 909 917 929 936 944
895 903 911 921 929 939
895 912 923 930 941
896 913 923 932 941
896 917 929 936 945
896 917 932 941
917 936
918 941
942
942
942
942

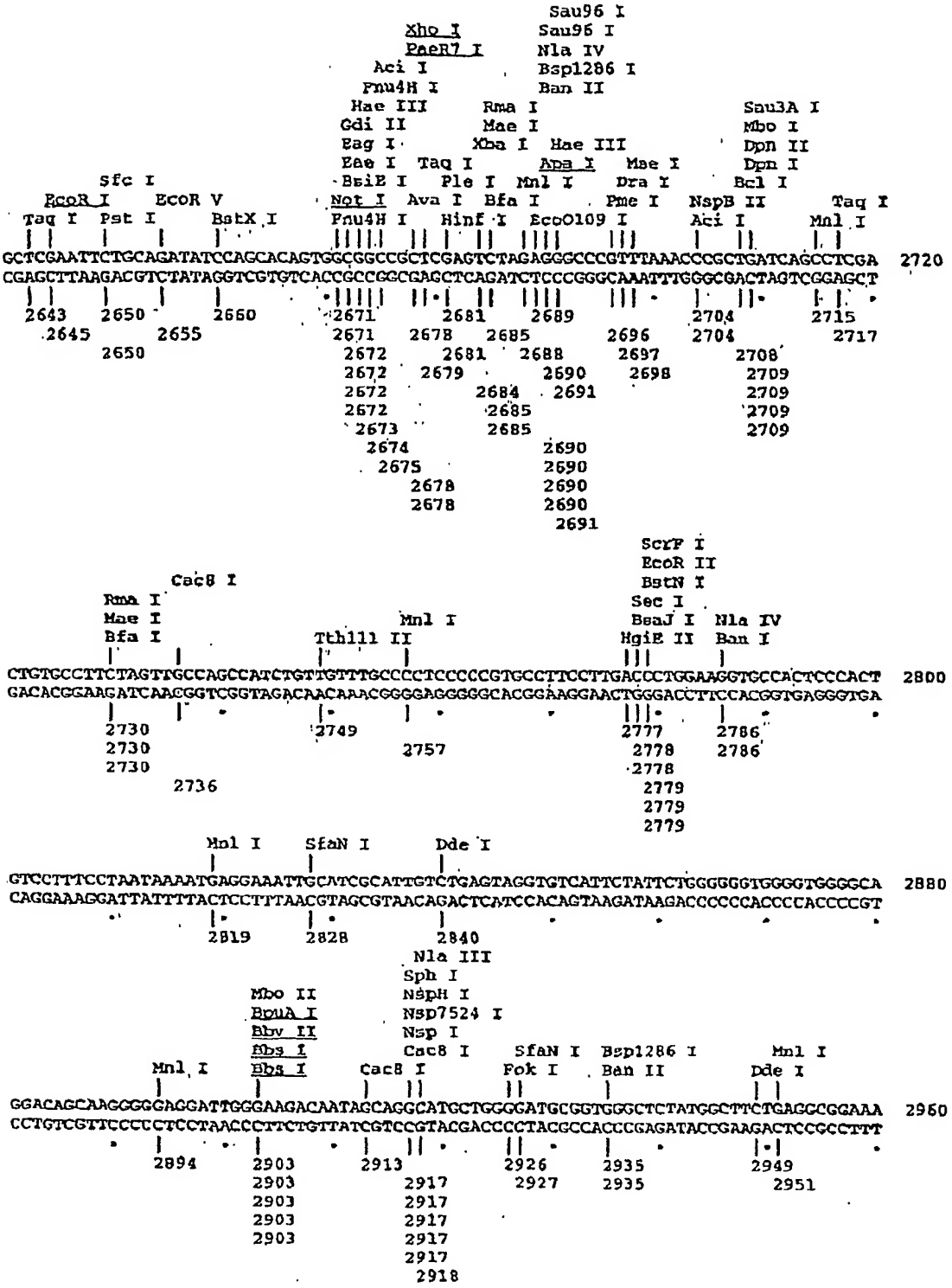

Tfi I
Taq I
Cla I
Sau3A I
Mbo I
Dpn II
BssS I
Hinf I
Ssp I
Mnl I
BspD I
Mae III
Mae III
Dpn I
Mnl I
CGTTACAGGAATATTGCAGTTACCTCGTGATCGATTCAAGAGGACATCATTCTTTCTTTGGGTAATTATCCTTTTCCAAA 1040
GCAATGTCCTTATAACGTCAATGGAGCACTAGCTAAGTTCTCCTGTAGTAAGAAAGAAACCCATTAATAGGAAAAGGTTT
962
970
979
983
984
989
990
993
989
989
989
990
991
993
Taq I
Rma I
Sal I
Mae I
HinC II
Bfa I
Eam1105 I
Xmn I
Fok I
Ple I
Hinf I
Acc I
Spe I
Mae III
GAACATTTTCCATCCCACTTGGAGTCATCCACAATAGCACATTACAGGTTAGTGATGTCGACAAACTAGTTTGTCGTGAC 1120
CTTGTAAAAGGTAGGGTGAACCTCAGTAGGTGTTATCGTGTAATGTCCAATCACTACAGCTGTTTGATCAAACAGCACTG
1041
1051
1062
1062
1066
1097
1097
1097
1098
1105
1106
1106
1106
1116
1118
Mme I
Sau3A I
Mbo I
Dpn II
Taq I
Tfi I
Fok I
Mun I
Dpn I
Hinf I
Mae II
AAACTGTCATCCACAAATCAATTGAGATCAGTTGGACTGAATCTCGAAGGGAATGGAGTGGCAACTGACGTGCCATCTGC 1200
TTTGACAGTAGGTGTTTAGTTAACTCTAGTCAACCTGACTTAGAGCTTCCCTTACCTCACCGTTGACTGCACGGTAGACG
1128
1139
1146
1146
1146
1146
1151
1159
1159
1164
1188
Msp I
Hpa II
BsaW I
Sau96 I
Ava II
HgiE II
Hph I
Eco57 I
Bsl I
Alu I
AACTAAAAGATGGGGCTTCAGGTCCGGTGTCCCACCAAAGGTGGTCAATTATGAAGCTGGTGAATGGGCTGAAAACTGCT 1280
TTGATTTTCTACCCCGAAGTCCAGGCCACAGGGTGGTTTCCACCAGTTAATACTTCGACCACTTACCCGACTTTTGACGA
1216
1221
1221
1223
1224
1224
1231
1234
1255
1259

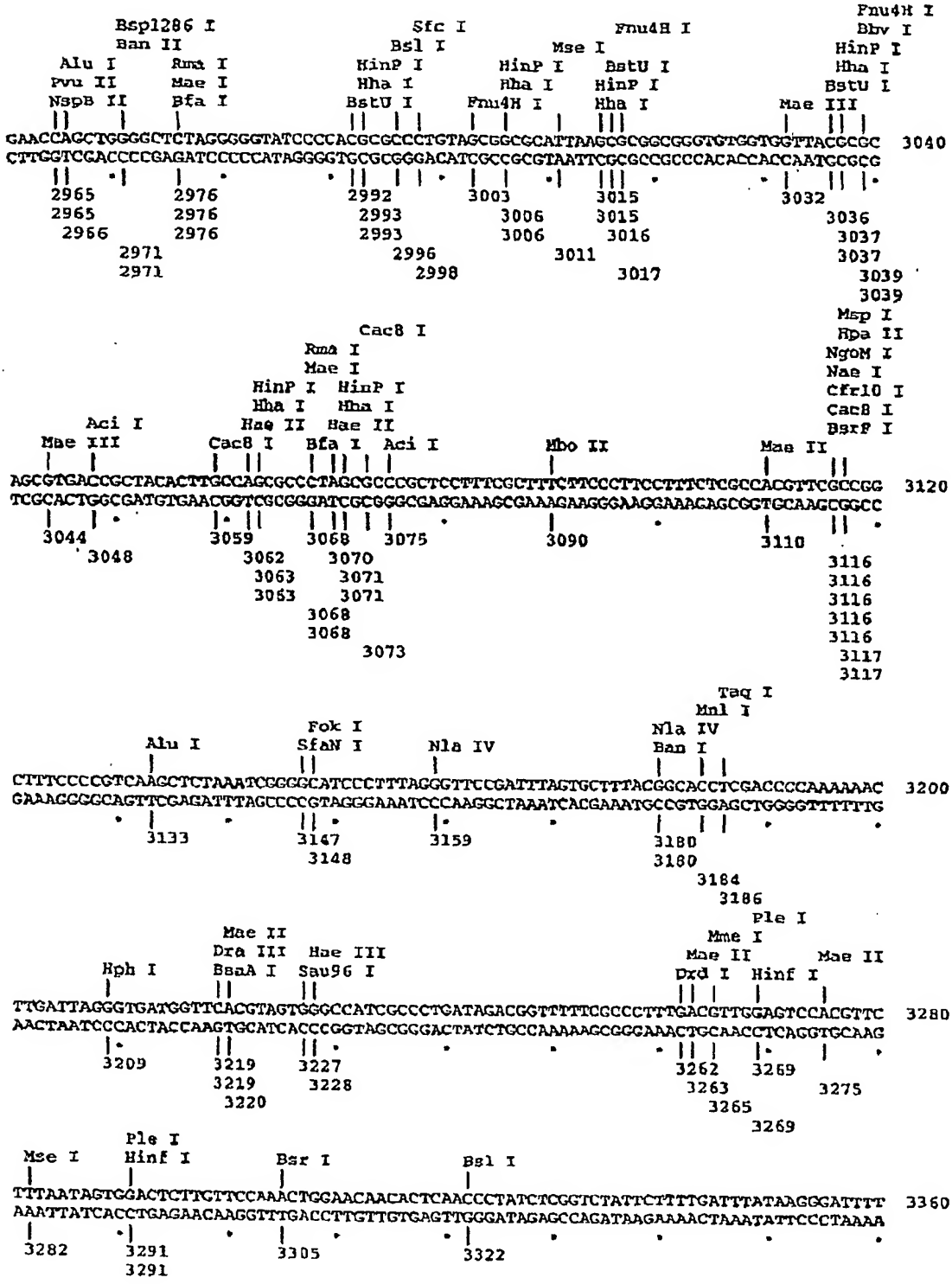

Cfr10 I
Nla IV
Ban I
Msp I
Hpa II
SerF I
Nci I
Bsl I
Bcn I
Msp I
Sec I
BsrF I
BsaJ I
Hpa II
HinP I
Hha I
Hae II
Fnu4H I
Tfi I
Acc I
Bbv I
Hinf I
ACAATCTTGAAATCAAAAAACCTGACGGGAGTGAGTGTCTACCAGCAGCGCCAGACGGGATTCGGGGCTTCCCCCGGTGC 1360
TGTTAGAACTTTAGTTTTTTGGACTGCCCTCACTCACAGATGGTCGTCGCGGTCTGCCCTAAGCCCCGAAGGGGGCCACG
1317 1325 1339 1352 1360
1325 1339 1352 1359
1327 1353 1360
1328 1353
1328 1353
1353
1354
1354
1356
1356
1359
HgiA I
Bsp1286 I
ApaL I
Alw21 I
Bsl I
Sau96 I
Ava II
Nla IV
BsmA I
Msp I
Hpa II
Bsl I
Mnl I
Mbo II
CGGTATGTGCACAAAGTATCAGGAACGGGACCGTGTGCCGGAGACTTTGCCTTCCATAAAGAGGGTGCTTTCTTCCTGTA 1440
GCCATACACGTGTTTCATAGTCCTTGCCCTGGCACACGGCCTCTGAAACGGAAGGTATTTCTCCCACGAAAGAAGGACAT
1367 1387 1398 1414 1421 1431
1367 1388 1398
1367 1388 1401
1367 1391
Taq I
Sau3A I
Mbo I
Dpn II
Dpn I
Mnl I
Sec I
BsaJ I
Eco57 I
Alu I
TGATCGACTTGCTTCCACAGTTATCTACCGAGGAACGACTTTCGCTGAAGGTGTCGTTGCATTTCTGATACTGCCCCAAG 1520
ACTAGCTGAACGAAGGTGTCAATAGATGGCTCCTTGCTGAAAGCGACTTCCACAGCAACGTAAAGACTATGACGGGGTTC
1442 1468 1485 1519
1442 1468
1442 1470
1442
1444
Bsl I
Sau96 I
Nla IV
Ava II
Rma I
Msp I
Hpa II
PpuM I
Mae I
Alu I
Xmn I
Eco57 I
Cfr10 I
Eco0109 I
Dde I
Mbo II
BsrF I
Mnl I
Bfa I
CTAAGAAGGACTTCTTCAGCTCACACCCCTTGAGAGAGCCGGTCAATGCAACGGAGGACCCGTCTAGTGGCTACTATTCT 1600
GATTCTTCCTGAAGAAGTCGAGTGTGGGGAACTCTCTCGGCCAGTTACGTTGCCTCCTGGGCAGATCACCGATGATAAGA
1521 1533 1558 1574 1584
1525 1534 1558 1575
1538 1559 1575 1584
1559 1576 1584
1576
1576
1580

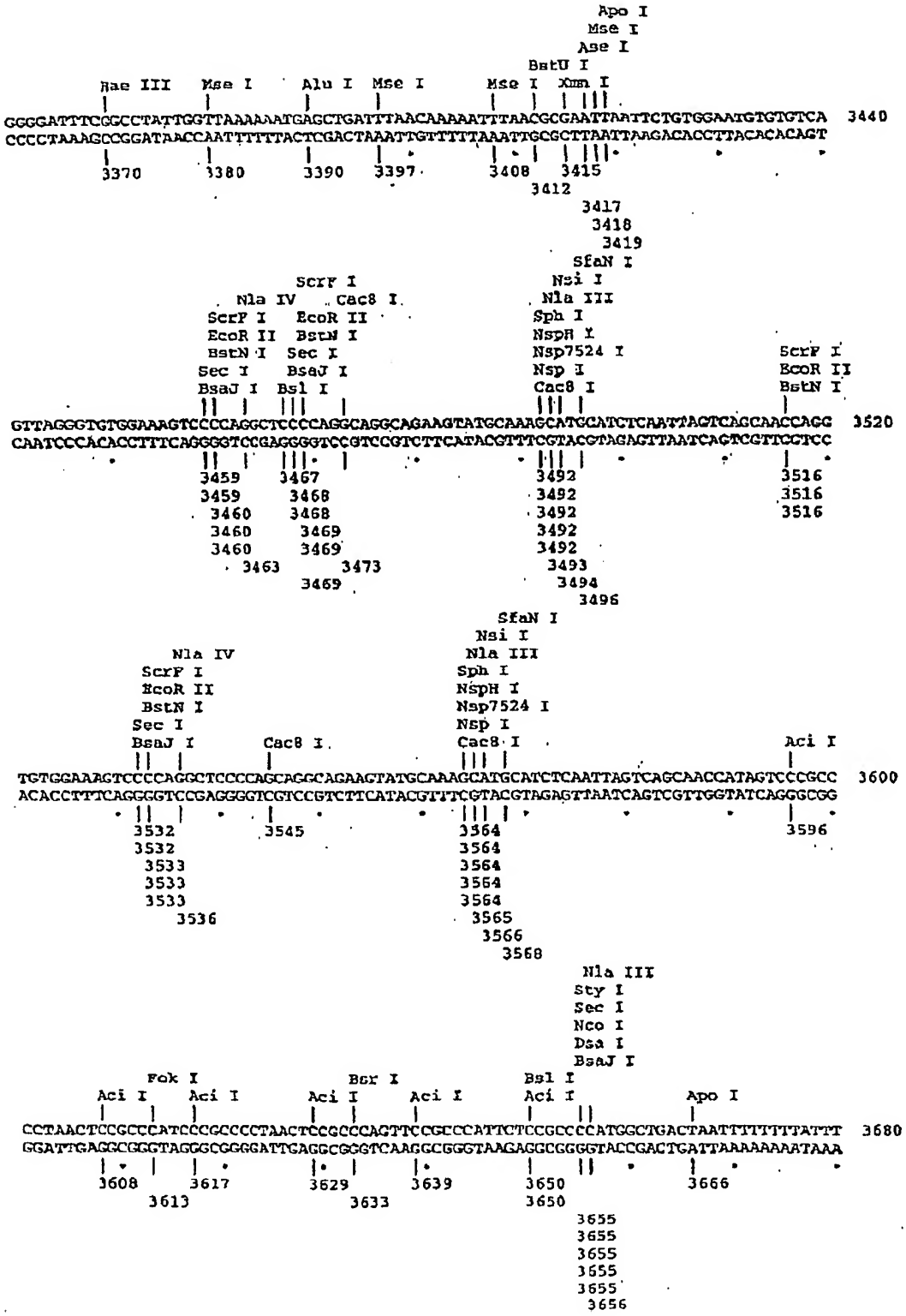

Msp I
Hpa II
Cfr10 I
BsrF I
BseW I
Age I
EcoR V
Nla IV
BsmA I
Sca I
Rsa I
HinC II
Mnl I
Taq I
Apo I
Mae II
Mme I
ACCACAATTAGATATCAGGCTACCGGTTTTGGAACCAATGAGACAGAGTACTTGTTCGAGGTTGACAATTTGACCTACGT 1680
TGGTGTTAATCTATAGTCCGATGGCCAAAACCTTGGTTACTCTGTCTCATGAACAAGCTCCAACTGTTAAACTGGATGCA
1611 1622 1631 1640 1648 1656 1666 1677
1622 1647 1658 1680
1622 1661
1622
1623
1623
Tfi I
Hinf I
Tfi I
Hinf I
Alu I
Pvu II
NspB II
AlwN I
Gsu I
BsmA I
CCAACTTGAATCAAGATTCACACCACAGTTTCTGCTCCAGCTGAATGAGACAATATATACAAGTGGGAAAAGGAGCAATA 1760
GGTTGAACTTAGTTCTAAGTGTGGTGTCAAAGACGAGGTCGACTTACTCTGTTATATATGTTCACCCTTTTCCTCGTTAT
1688 1706 1715 1727
1688
1695
1695
1718
1718
1719
Sec I
Dsa I
BsaJ I
Apo I
HinC II
Hae III
Sau96 I
CCACGGGAAAACTAATTTGGAAGGTCAACCCCGAAATTGATACAACAATCGGGGAGTGGGCCTTCTGGGAAACTAAAAAA 1840
GGTGCCCTTTTGATTAAACCTTCCAGTTGGGGCTTTAACTATGTTGTTAGCCCCTCACCCGGAAGACCCTTTGATTTTTT
1761 1773 1784 1818
1761 1819
1761
codon 308
codon 490
Rma I
Mae I
Bfa I
Mbo II
Ear I
Mnl I
Rma I
Mae I
Bfa I
Xba I
Sau3A I
Mbo I
Dpn II
Dpn I
Bcl I
Mbo II
Taq I
Ear I
AACCTCACTAGAAAAATTCGCAGTGAAGAGTTGTCTTTCTCTAGAGCAGGACTGATCACAGGCGGGAGAAGAACTCGAAG 1920
TTGGAGTGATCTTTTTAAGCGTCACTTCTCAACAGAAAGAGATCTCGTCCTGACTAGTGTCCGCCCTCTTCTTGAGCTTC
1843 1865 1880 1893 1908 1917
1848 1865 1881 1894 1915
1848 1881 1894 1917
1881 1894
1894
Mun I
Apo I
Bsr I
Dde I
Fok I
Fnu4H I
Bbv I
AGAAGCAATTGTCAATGCTCAACCCAAATGCAACCCTAATTTACATTACTGGACTACTCAGGATGAAGGTGCTGCAATCG 2000
TCTTCGTTAACAGTTACGAGTTGGGTTTACGTTGGGATTAAATGTAATGACCTGATGAGTCCTACTTCCACGACGTTAGC
1926 1957 1968 1977 1981 1991
1991
ScrF I
EcoR II
BstN I
Fnu4H I
Bbv I
Hae III
Hae I
Cac8 I
Hae III
Mnl I
Sec I
Bsr I
Bsl I
Sau96 I
BsaJ I
Mnl I
Mse I
GACTGGCCTGGATACCATATTTCCGGCCAGCAGCCGAGGGAATTTACATAGAGGGGCTAATGCACAATCAAGATGGTTTA 2080
CTGACCGGACCTATGGTATAAAGCCCGGTCGTCGGCTCCCTTAAATGTATCTCCCCGATTACGTGTTAGTTCTACCAAAT
2002 2015 2024 2034 2051 2078
2004 2025 2034
2005 2026 2036
2007 2030
2007 2030
2007

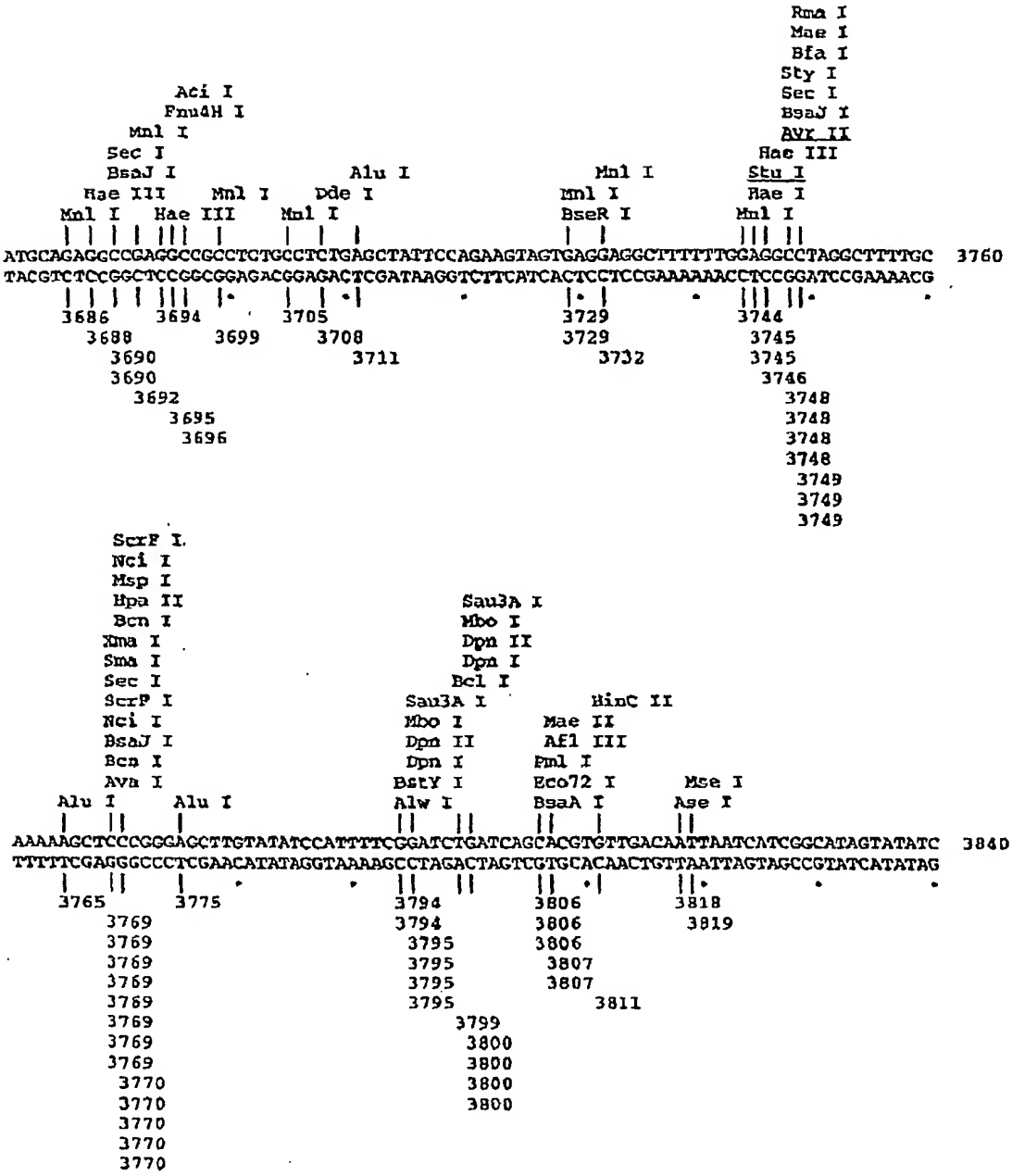

Hae III
Msc I
Hae I
Cac8 I
Alu I
Pvu II
NspB II
Ple I
Hinf I
Mbo II
Ear I
BsmA I
Eae I
BsmA I
Alu I
Pde I
Dde I
Alu I
ATCTGTGGGTTGAGACAGCTGGCCAACGAGACGACTCAAGCTCTTCAACTGTTCCTGAGAGCCACAACTGAGCTACGCAC 2160
TAGACACCCAACTCTGTCGACCGGTTGCTCTGCTGAGTTCGAGAAGTTGACAAGGACTCTCGGTGTTGACTCGATGCGTG
2092
2100
2108
2119
2135
2148
2096
2113
2121
2151
2096
2113
2122
2097
2098
2100
2100
2101
Fnu4H I
Bbv I
Sfc I
Pst I
Fnu4H I
Nsp I
PflM I
Nla III
NspH I
Nsp7524 I
Msp I
Hpa II
BsaW I
Sau96 I
Ava II
Kpn I
Mun I
Bbv I
Fnu4H I
Bsl I
Nla IV
CTTTTCAATCCTCAACCGTAAGGCAATTGATTTCTTGCTGCAGCGATGGGGCGGCACATGCCACATTCTGGGACCGGACT 2240
GAAAAGTTAGGAGTTGGCATTCCGTTAACTAAAGAACGACGTCGCTACCCCGCCGTGTACGGTGTAAGACCCTGGCCTGA
2170
2184
2197
2197
2198
2198
2200
2200
2211
2216
2216
2216
2217
2221
2221
2230
2231
2231
2233
2234
2234
Sau3A I
Mbo I
Dpn II
Dpn I
Bcl I
Sau96 I
Ava II
BstX I
Taq I
Nla III
BspH I
Nla III
GCTGTATCGAACCACATGATTGGACCAAGAACATAACAGACAAAAATTGATCAGATTATTCATGATTTTGTTGATAAAACC 2320
CGACATAGCTTGGTGTACTAACCTGGTTCTTGTATTGTCTGTTTTAACTAGTCTAATAAGTACTAAAACAACTATTTTGG
2247
2255
2252
2262
2262
2287
2288
2288
2288
2288
2299
2300
Sec I
ScrF I
EcoR II
BstN I
BsaJ I
Sau96 I
Ava II
Msp I
Hpa II
Bsl I
BspE I
BsaW I
Mun I
Fok I
BsmA I
BspM I
Msp I
Hpa II
CfrI0 I
BsrF I
Mae III
CTTCCGGACCAGGGGGACAATGACAATTGGTGGACAGGATGGAGACAATGGATACCGGCAGGTATTGGAGTTACAGGCGT 2400
GAAGGCCTGGTCCCCCTGTTACTGTTAACCACCTGTCCTACCTCTGTTACCTATGGCCGTCCATAACCTCAATGTCCGCA
2323
2323
2324
2324
2324
2326
2326
2329
2329
2329
2329
2329
2344
2357
2362
2374
2374
2375
2375
2378
2390

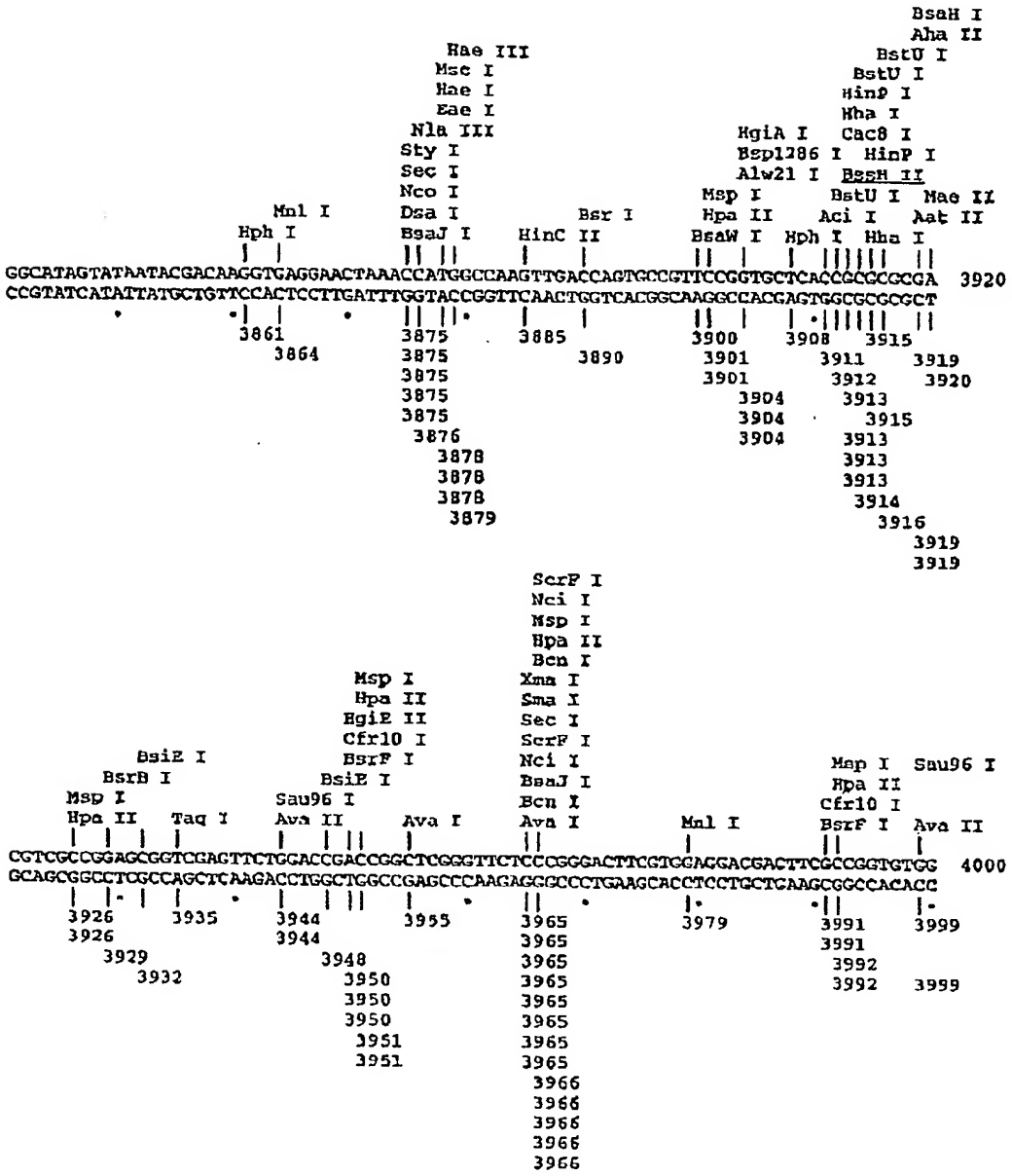

GP Δ309-489
Stop
(TAG)
Eco57 I
Mbo II
Nla III
Dde I
Esp I
Alu I
Bpu1102 I
TATAATTGCAGTTATCGCTTTATTCTGTATATGCAAATTTGTCTTTTACTTTTTCTTCAGATTGCTTCATGGAAAAGCTC 2480
ATATTAACGTCAATAGCGAAATAAGACATATACGTTTAAACAGAAAATCAAAAAGAAGTCTAACGAAGTACCTTTTCGAG
2454
2455
2468
2477
2476
2477
2478
Mse I
ScrF I
RcoR II
BstN I
Mnl I
Dde I
Tfi I
Hinf I
AGCCTCAAATCAATGAAACCAGGATTTAATTATATGGATTACTTGAATCTAAGATTACTTGACAAATGATAATATAATAC 2560
TCGGAGTTTAGTTACTTTGGTCCTAAATTAATATACCTAATGAACTTAGATTCTAATGAACTGTTTACTATTATATTATG
2483
2499
2499
2499
2506
2525
2525
2529
Sac I
BgiA I
Ecl136 II
Bsp1286 I
Ban II
Rsa I
Nla IV
Kpn I
Ban I
Asp718
Acc65 I
Alu I
Gsu I
Bpm I
Bsr I
Alu I
Mse I
Dra I
Tfi I
Hinf I
Mse I
Dra I
Mse I
Mnl I
Alw21 I
ACTGGAGCTTTAAACATAGCCAATGTGATTCTAACTCCTTTAAACTCACAGTTAATCATAAACAAGGTTTGAGGTACCGA 2640
TGACCTCGAAATTTGTATCGGTTACACTAAGATTGAGGAAATTTGAGTGTCAATTAGTATTTGTTCCAAACTCCATGGCT
2561
2562
2562
2566
2569
2570
2587
2587
2599
2600
2612
2631
2633
2633
2633
2633
2633
2634
2639
2640
2639
2639
2639
2639
2639

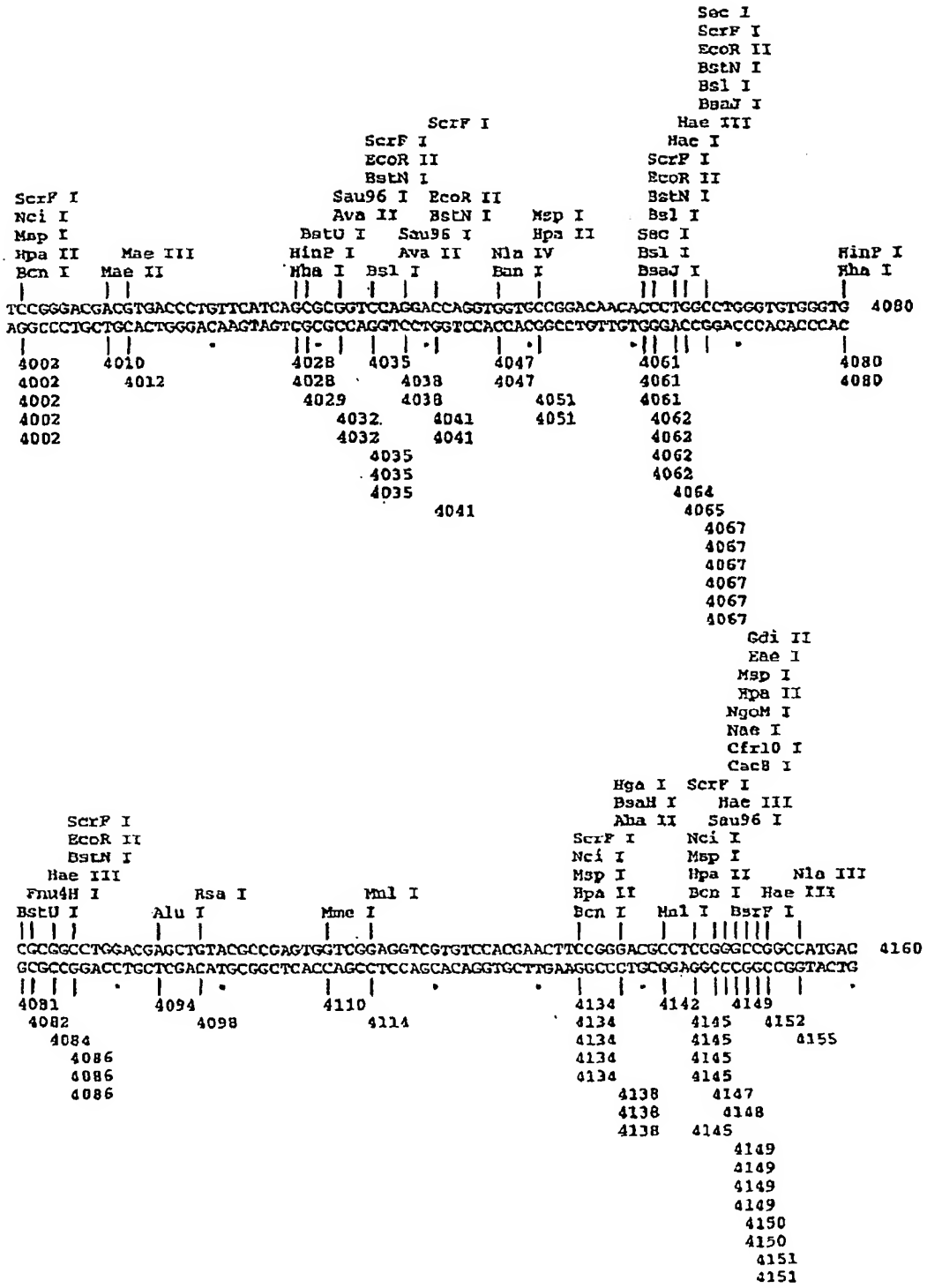

Sau96 I
Xho I
Sau96 I
PaeR7 I
Nla IV
Aci I
Bsp1286 I
Fnu4H I
Ban II
Hae III
Rma I
Sau3A I
Gdi II
Mae I
Mbo I
Eag I
Xba I
Hae III
Dpn II
Eae I
Taq I
Apa I
Mse I
Dpn I
Sfc I
BsiE I
Ple I
Mnl I
Dra I
Bcl I
EcoR I
EcoR V
Not I
Ava I
Bfa I
Pme I
NspB II
Taq I
Taq I
Pst I
BstX I
Fnu4H I
Hinf I
Eco0109 I
Aci I
Mnl I
GCTCGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGA 2720
CGAGCTTAAGACGTCTATAGGTCGTGTCACCGCCGGCGAGCTCAGATCTCCCGGGCAAATTTGGGCGACTAGTCGGAGCT
2643 2650 2660 2671 2681 2689 2704 2715
2645 2655 2671 2678 2685 2696 2704 2717
2650 2672 2681 2688 2697 2708
2672 2679 2690 2698 2709
2672 2684 2691 2709
2672 2685 2709
2673 2685 2709
2674 2690
2675 2690
2678 2690
2678 2690
2691
ScrF I
EcoR II
BstN I
Cac8 I
Rma I
Sec I
Mae I
Mnl I
BsaJ I
Nla IV
Bfa I
Tth111 II
HgiE II
Ban I
CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACT 2800
GACACGGAAGATCAACGGTCGGTAGACAACAAACGGGGAGGGGGCACGGAAGGAACTGGGACCTTCCACGGTGAGGGTGA
2730 2749 2757 2777 2786
2730 2778 2786
2730 2778
2736 2779
2779
2779
Mnl I
SfaN I
Dde I
GTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCA 2880
CAGGAAAGGATTATTTTACTCCTTTAACGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCCCCACCCCACCCCGT
2819 2828 2840
Nla III
Sph I
Mbo II
NsiH I
BpuA I
Nsp7524 I
Bbv II
Nsp I
Bbs I
Cac8 I
SfaN I
Bsp1286 I
Mnl I
Mnl I
Bbs I
Cac8 I
Fok I
Ban II
Dde I
GGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAA 2960
CCTGTCGTTCCCCCTCCTAACCCTTCTGTTATCGTCCGTACGACCCCTACGCCACCCGAGATACCGAAGACTCCGCCTTT
2894 2903 2913 2926 2935 2949
2903 2917 2927 2935 2951
2903 2917
2903 2917
2903 2917
2917
2918

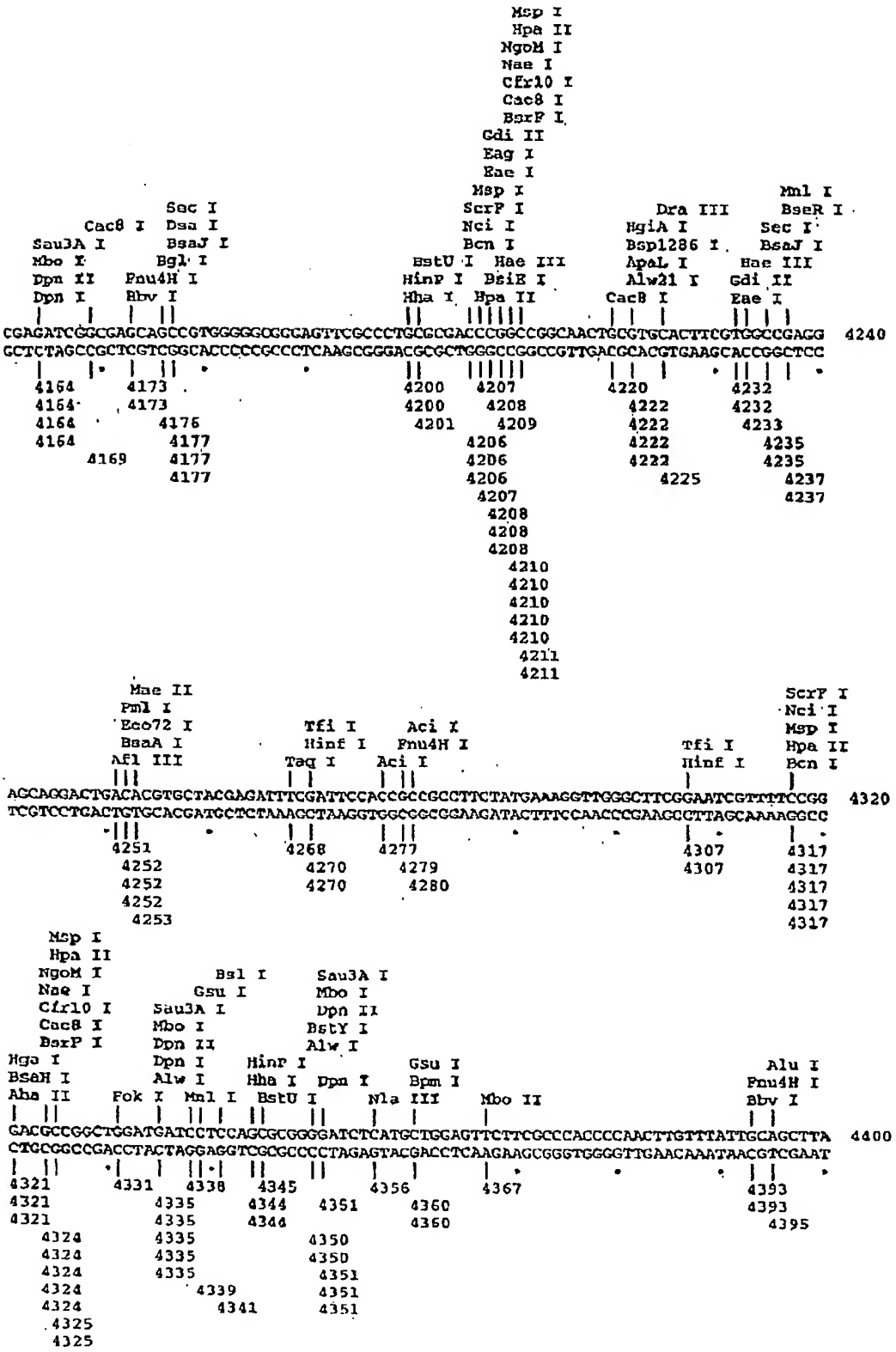

Bsp1286 I
Ban II
Alu I
Pvu II
NspB II
Rma I
Mae I
Bfa I
Sfc I
Bsl I
HinP I
Hha I
BstU I
Mse I
HinP I
Hha I
Fnu4H I
Fnu4H I
BstU I
HinP I
Hha I
Fnu4H I
Bbv I
HinP I
Hha I
BstU I
Mae III
GAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGC 3040
CTTGGTCGACCCCGAGATCCCCCATAGGGGTGCGCGGGACATCGCCGCGTAATTCGCGCCGCCCACACCACCAATGCGCG
2965
2965
2966
2976
2976
2976
2971
2971
2992
2993
2993
2996
2998
3003
3006
3006
3011
3015
3015
3016
3017
3032
3036
3037
3037
3039
3039
Msp I
Hpa II
NgoM I
Nae I
Cfr10 I
Cac8 I
BsrF I
Cac8 I
Rma I
Mae I
HinP I
Hha I
HinP I
Hha I
Aci I
Mae III
Hae II
Hae II
Cac8 I
Bfa I
Aci I
Mbo II
Mae II
AGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGG 3120
TCGCACTGGCGATGTGAACGGTCGCGGGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAAGCGGCC
3044
3048
3059
3062
3063
3063
3068
3068
3068
3070
3071
3071
3073
3075
3090
3110
3116
3116
3116
3116
3116
3117
3117
Taq I
Mnl I
Fok I
SfaN I
Alu I
Nla IV
Nla IV
Ban I
CTTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAAC 3200
GAAAGGGGCAGTTCGAGATTTAGCCCCGTAGGGAAATCCCAAGGCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTG
3133
3147
3148
3159
3180
3180
3184
3186
Ple I
Mae II
Dra III
BsaA I
Hae III
Sau96 I
Mme I
Mae II
Dxd I
Hinf I
Mae II
Hph I
TTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTC 3280
AACTAATCCCACTACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACTGCAACCTCAGGTGCAAG
3209
3219
3219
3220
3227
3228
3262
3263
3265
3269
3269
3275
Ple I
Mse I
Hinf I
Bsr I
Bsl I
TTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTT 3360
AAATTATCACCTGAGAACAAGGTTTGACCTTGTTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCCTAAAA
3282
3291
3291
3305
3322

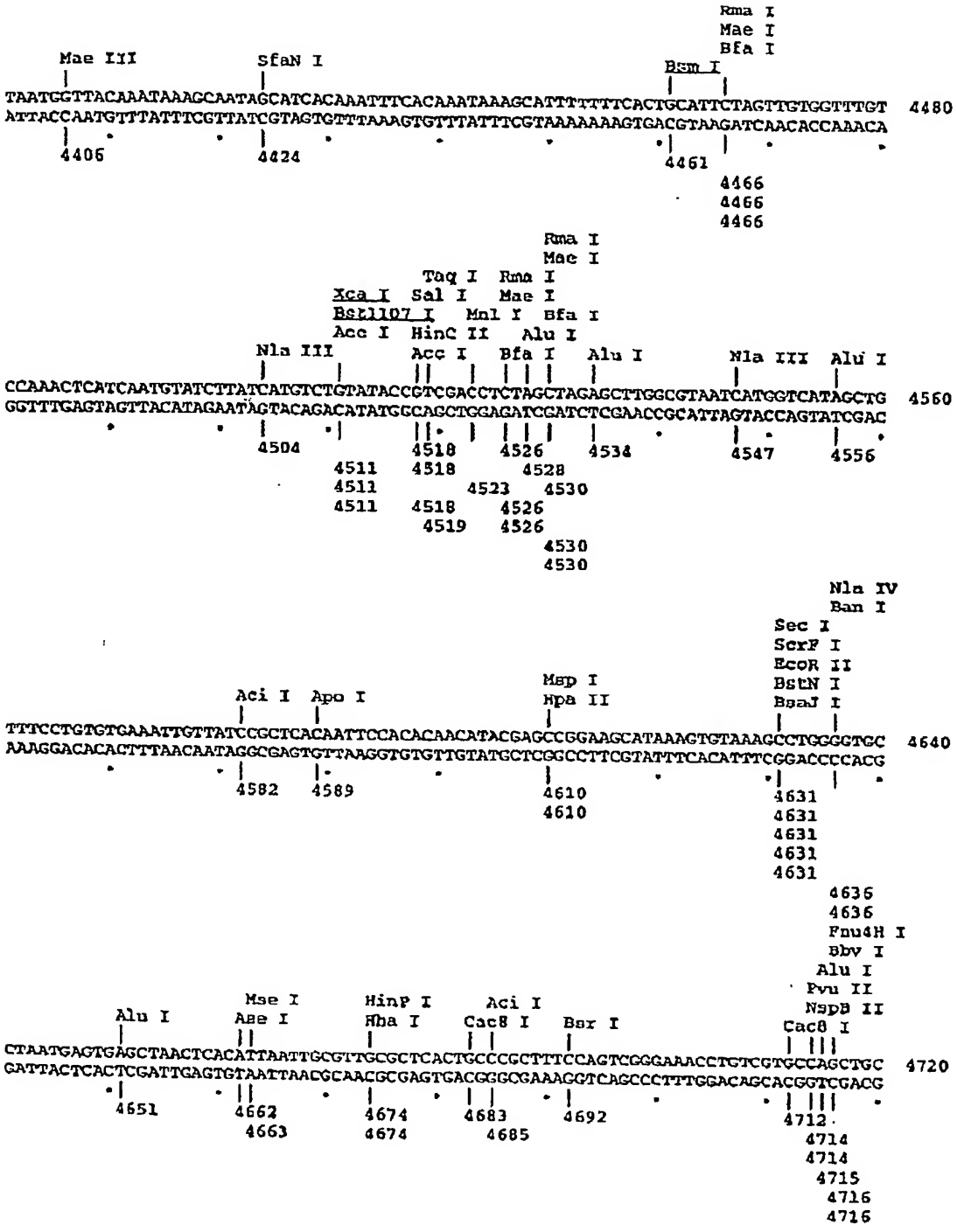

Apo I
Mse I
Ase I
BstU I
Rae III
Mse I
Alu I
Mse I
Mse I
Xmn I
GGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCA 3440
CCCCTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTGTTTTTAAATTGCGCTTAATTAAGACACCTTACACACAGT
3370
3380
3390
3397
3408
3415
3412
3417
3418
3419
SfaN I
Nsi I
ScrF I
Nla IV
Cac8 I
Nla III
ScrF I
EcoR II
Sph I
EcoR II
BstN I
NspH I
BstN I
Sec I
Nsp7524 I
ScrF I
Sec I
BsaJ I
Nsp I
EcoR II
BsaJ I
Bsl I
Cac8 I
BstN I
GTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGG 3520
CAATCCCACACCTTTCAGGGGTCCGAGGGGTCCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTCCTCC
3459
3467
3492
3516
3459
3468
3492
3516
3460
3468
3492
3516
3460
3469
3492
3460
3469
3492
3463
3473
3493
3469
3494
3496
SfaN I
Nsi I
Nla IV
Nla III
ScrF I
Sph I
EcoR II
NspH I
BstN I
Nsp7524 I
Sec I
Nsp I
BsaJ I
Cac8 I
Cac8 I
Aci I
TGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCC 3600
ACACCTTTCAGGGGTCCGAGGGGTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGG
3532
3545
3564
3596
3532
3564
3533
3564
3533
3564
3533
3564
3536
3565
3566
3568
Nla III
Sty I
Sec I
Nco I
Dsa I
BsaJ I
Fok I
Bsr I
Bsl I
Aci I
Aci I
Aci I
Aci I
Aci I
Apo I
CCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTT 3680
GGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAAAAAAAATAAA
3608
3617
3629
3639
3650
3666
3613
3633
3650
3655
3655
3655
3655
3655
3656

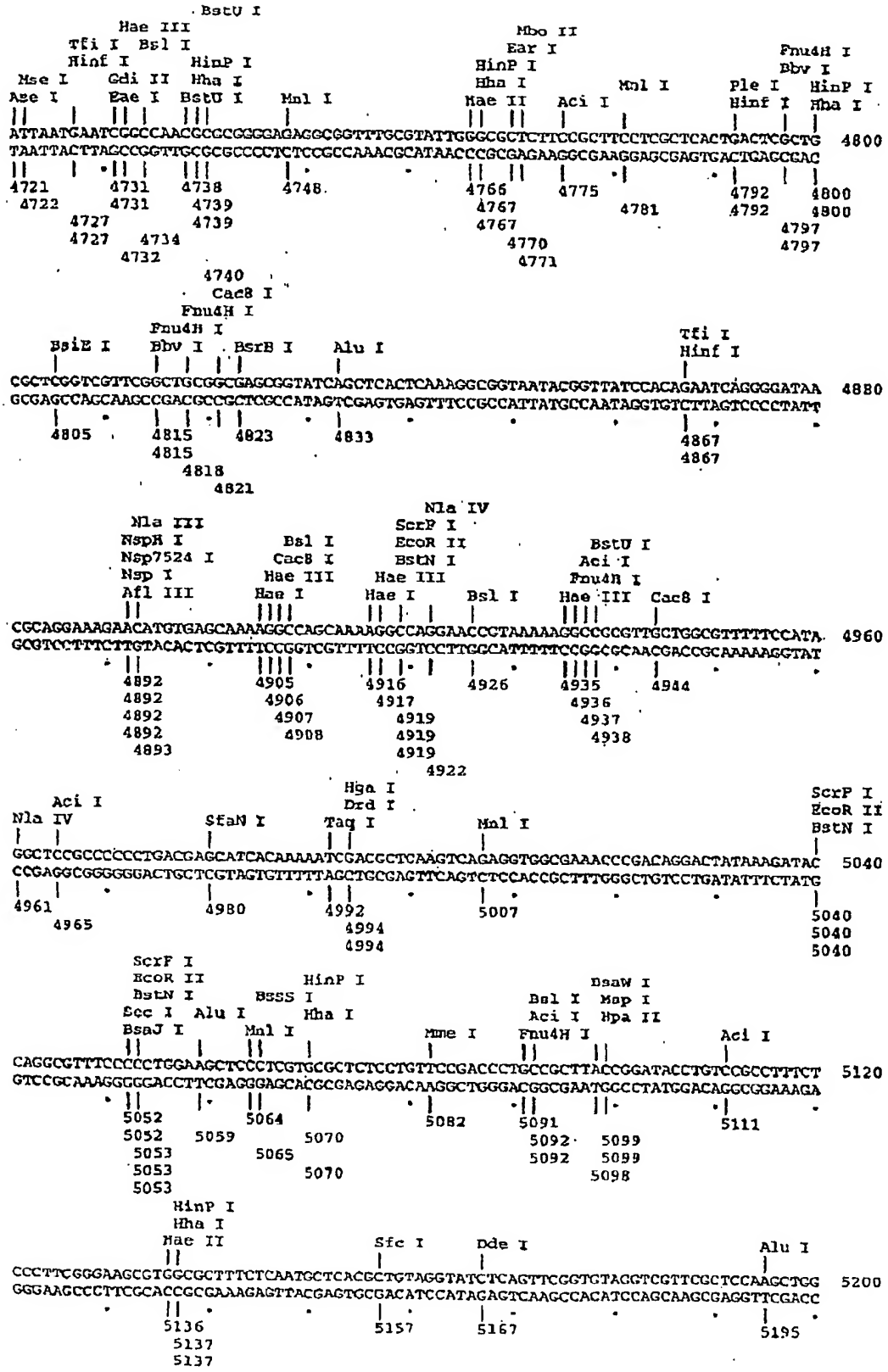

Rma I
Mae I
Bfa I
Sty I
Sec I
BsaJ I
AvrII
Hae III
Stu I
Hae I
Mnl I
Aci I
Fnu4H I
Mnl I
Sec I
BsaJ I
Hae III
Mnl I
Hae III
Mnl I
Alu I
Dde I
Mnl I
BseR I
ATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGC 3760
TACGTCTCCGGCTCCGGCGGAGACGGAGACTCGATAAGGTCTTCATCACTCCTCCGAAAAAACCTCCGGATCCGAAAACG
3686
3688
3690
3690
3692
3694
3695
3696
3699
3705
3708
3711
3729
3729
3732
3744
3745
3745
3746
3748
3748
3748
3748
3749
3749
3749
ScrF I
Nci I
Msp I
Hpa II
Bcn I
Xma I
Sma I
Sec I
ScrF I
Nci I
BsaJ I
Bcn I
Ava I
Sau3A I
Mbo I
Dpn II
Dpn I
Bcl I
Sau3A I
Mbo I
Dpn II
Dpn I
BstY I
Alw I
HinC II
Mae II
AfI III
Pml I
Eco72 I
BsaA I
Mse I
Ase I
Alu I
Alu I
AAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAGCACGTGTTGACAATTAATCATCGGCATAGTATATC 3840
TTTTTCGAGGGCCCTCGAACATATAGGTAAAAGCCTAGACTAGTCGTGCACAACTGTTAATTAGTAGCCGTATCATATAG
3765
3769
3769
3769
3769
3769
3769
3769
3769
3770
3770
3770
3770
3770
3775
3794
3794
3795
3795
3795
3795
3799
3800
3800
3800
3800
3806
3806
3806
3807
3807
3811
3818
3819

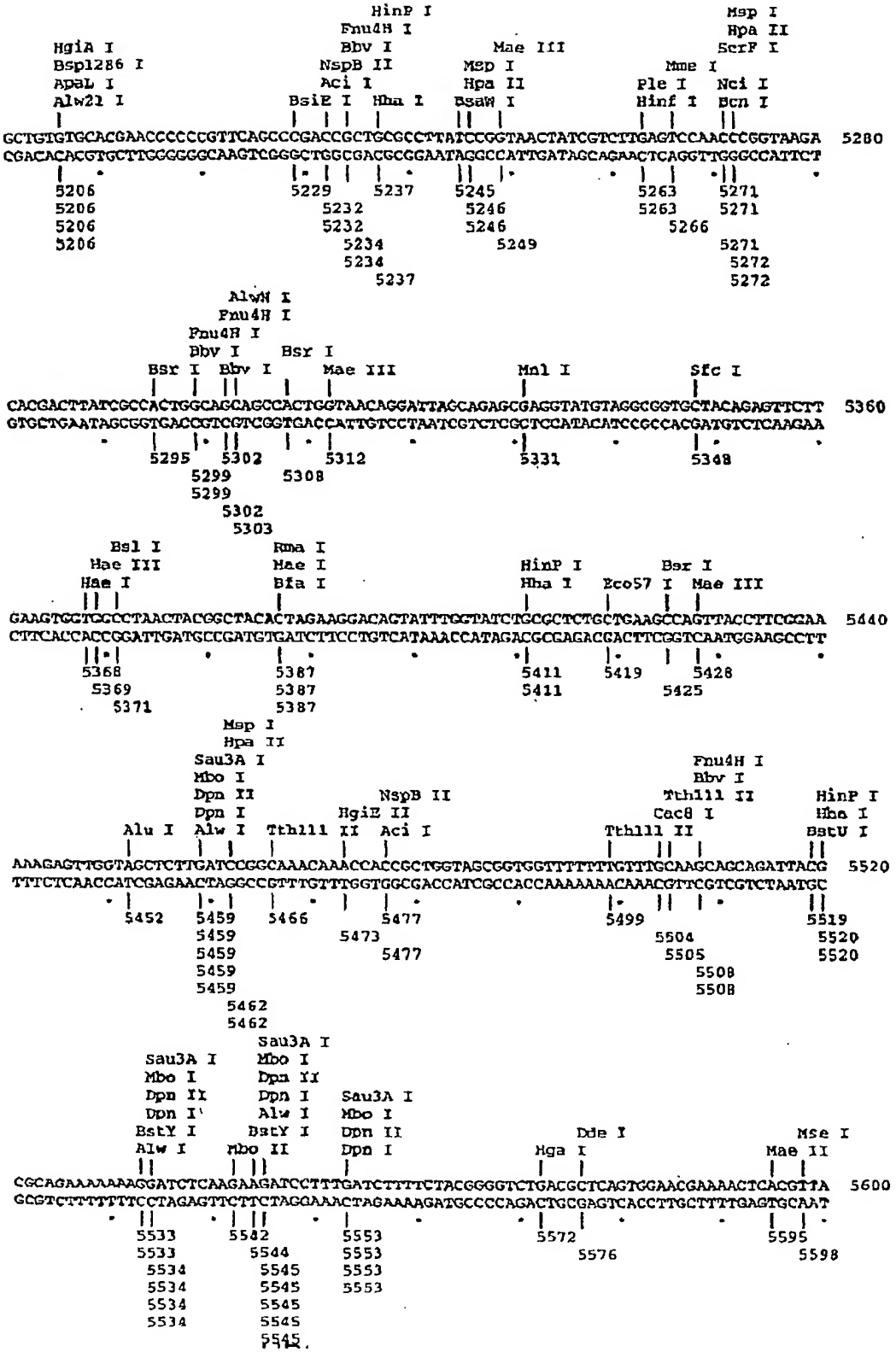

BsaH I
Aha II
BstU I
BstU I
HinP I
Hha I
Cac8 I
HinP I
BssH II
Mae II
Aat II
Hha I
Rae III
Msc I
Hae I
Eae I
Nla III
Sty I
Sec I
Nco I
Dsa I
BsaJ I
Mnl I
Hph I
Bsr I
HinC II
HgiA I
Bsp1286 I
Alw21 I
Msp I
Hpa II
BsaW I
Hph I
Aci I
GGCATAGTATAATACGACAAGGTGAGGAACTAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGA 3920
CCGTATCATATTATGCTGTTCCACTCCTTGATTTGGTACCGGTTCAACTGGTCACGGCAAGGCCACGAGTGGCGCGCGCT
3861
3864
3875
3875
3875
3875
3875
3876
3878
3878
3878
3879
3885
3890
3900
3901
3901
3904
3904
3904
3908
3911
3912
3913
3915
3913
3913
3913
3914
3916
3915
3919
3919
3919
3920
ScrF I
Nci I
Msp I
Hpa II
Ben I
Xma I
Sma I
Sec I
ScrF I
Nci I
BsaJ I
Ben I
Ava I
Msp I
Hpa II
HgiE II
Cfr10 I
BsrF I
BsiE I
Sau96 I
Ava II
BsiE I
BsrB I
Msp I
Hpa II
Taq I
Ava I
Mnl I
Msp I
Hpa II
Cfr10 I
BsrF I
Sau96 I
Ava II
CGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGG 4000
GCAGCGGCCTCGCCAGCTCAAGACCTGGCTGGCCGAGCCCAAGAGGGCCCTGAAGCACCTCCTGCTGAAGCGGCCACACC
3926
3926
3929
3932
3935
3944
3944
3948
3950
3950
3950
3951
3951
3955
3965
3965
3965
3965
3965
3965
3965
3966
3966
3966
3966
3966
3979
3991
3991
3992
3992
3999
3999

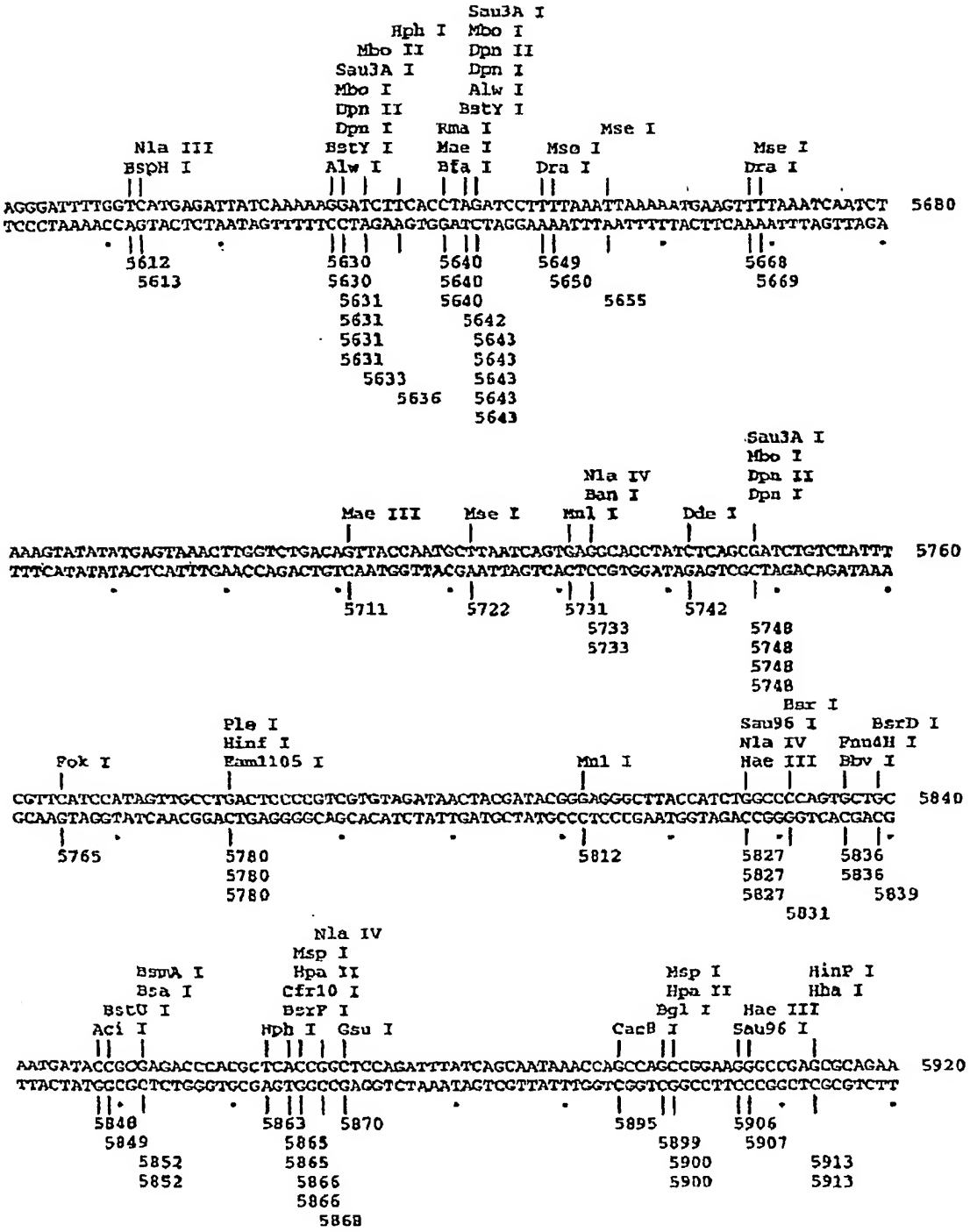

Sec I
ScrF I
EcoR II
BstN I
Bsl I
BsaJ I
Hae III
Hae I
ScrF I
EcoR II
BstN I
Bsl I
Sec I
BsaJ I
ScrF I
ScrF I
EcoR II
BstN I
Sau96 I
Ava II
EcoR II
BstN I
Msp I
Hpa II
BstU I
Sau96 I
Ava II
Nla IV
Ban I
HinP I
Hha I
Bsl I
ScrF I
Nci I
Msp I
Hpa II
Bcn I
Mae III
Mae II
HinP I
Hha I
TCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGTGTGGGTG 4080
AGGCCCTGCTGCACTGGGACAAGTAGTCGCGCCAGGTCCTGGTCCACCACGGCCTGTTGTGGGACCGGACCCACACCCAC
4002
4002
4002
4002
4002
4010
4012
4028
4028
4029
4032
4032
4035
4035
4035
4035
4038
4038
4041
4041
4041
4047
4047
4051
4051
4061
4061
4061
4062
4062
4062
4062
4064
4065
4067
4067
4067
4067
4067
4067
4080
4080
Gdi II
Eae I
Msp I
Hpa II
NgoM I
Nae I
Cfr10 I
CacB I
Hga I
BsaH I
Aha II
ScrF I
Hae III
Sau96 I
ScrF I
Nci I
Msp I
Hpa II
Bcn I
Nci I
Msp I
Hpa II
Bcn I
Mnl I
Hae III
BsrF I
Nla III
ScrF I
EcoR II
BstN I
Hae III
Fnu4H I
BstU I
Alu I
Rsa I
Mme I
Mnl I
CGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGAC 4160
GCGCCGGACCTGCTCGACATGCGGCTCACCAGCCTCCAGCACAGGTGCTTGAAGGCCCTGCGGAGGCCCGGCCGGTACTG
4081
4082
4084
4086
4086
4086
4094
4098
4110
4114
4134
4134
4134
4134
4134
4138
4138
4138
4142
4145
4145
4145
4145
4147
4148
4145
4149
4149
4149
4149
4150
4150
4151
4151
4149
4152
4155

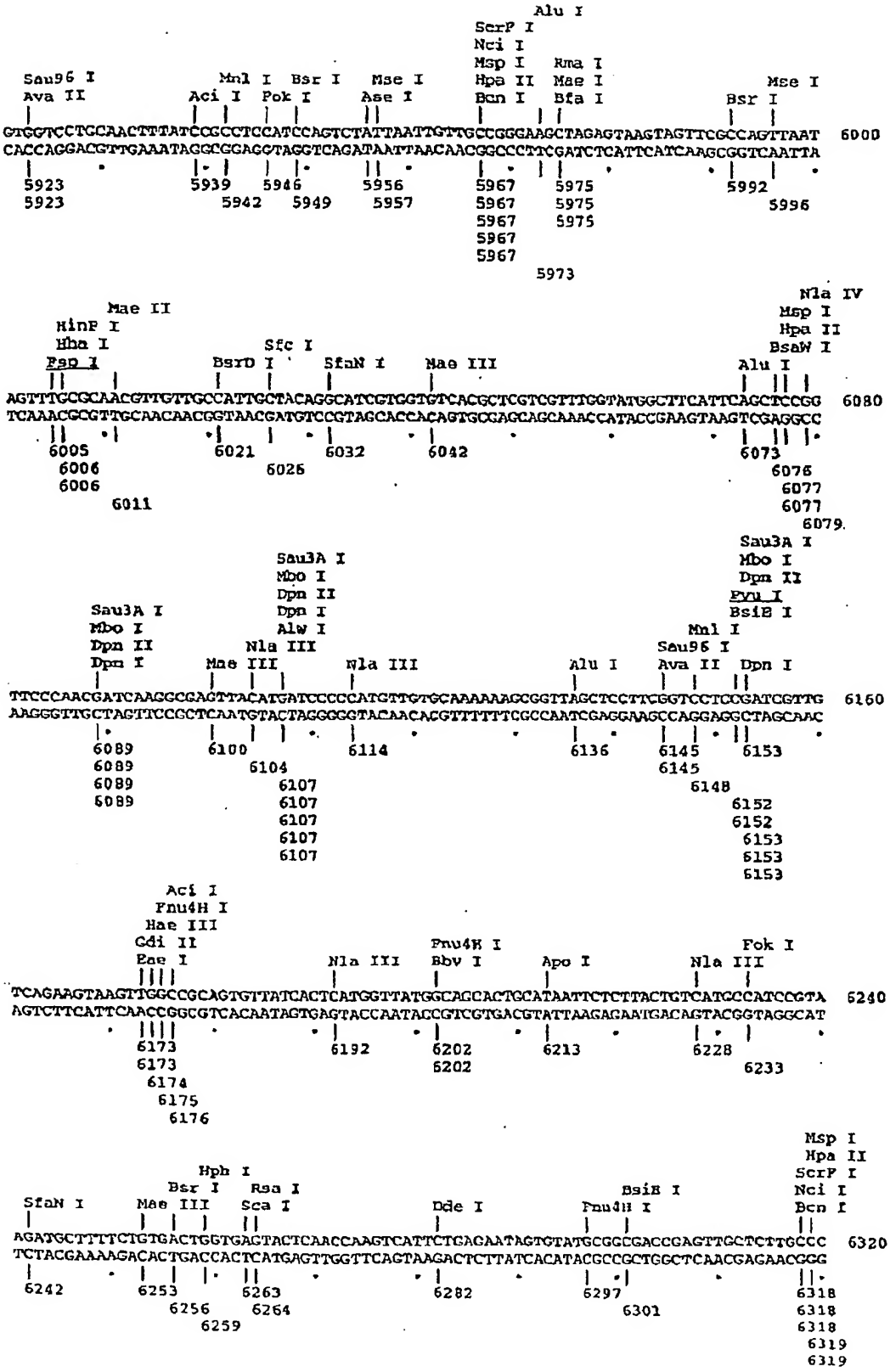

Msp I
Hpa II
NgoM I
Nae I
Cfr10 I
Cac8 I
BsrF I
Gdi II
Eag I
Eae I
Msp I
ScrF I
Nci I
Bcn I
BstU I
Hae III
HinP I
Hha I
BsiE I
Bpa II
Sec I
Dsa I
BsaJ I
Bgl I
Pnu4H I
Bbv I
Cac8 I
Sau3A I
Mbo I
Dpn II
Dpn I
Dra III
HgiA I
Bsp1286 I
ApaL I
Alw21 I
Cac8 I
Mnl I
BseR I
Sec I
BsaJ I
Hae III
Gdi II
Eae I
CGAGATCGGCGAGCAGCCGTGGGGGCGGCAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGG 4240
GCTCTAGCCGCTCGTCGGCACCCCCGCCCTCAAGCGGGACGCGCTGGGCCGGCCGTTGACGCACGTGAAGCACCGGCTCC
4164
4164
4164
4164
4169
4173
4173
4176
4177
4177
4177
4200
4200
4201
4206
4206
4206
4207
4208
4208
4208
4210
4210
4210
4210
4210
4211
4211
4207
4208
4209
4220
4222
4222
4222
4222
4225
4232
4232
4233
4235
4235
4237
4237
Mae II
Pml I
Eco72 I
BsaA I
Afl III
Tfi I
Hinf I
Taq I
Aci I
Fnu4H I
Aci I
Tfi I
Hinf I
ScrF I
Nci I
Msp I
Hpa II
Bcn I
AGCAGGACTGACACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGG 4320
TCGTCCTGACTGTGCACGATGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCC
4251
4252
4252
4252
4253
4268
4270
4270
4277
4279
4280
4307
4307
4317
4317
4317
4317
4317
Msp I
Hpa II
NgoM I
Nae I
Cfr10 I
Cac8 I
BsrF I
Bsl I
Gsu I
Sau3A I
Mbo I
Dpn II
Sau3A I
Mbo I
Dpn II
BstY I
Alw I
Hga I
BseH I
Aha II
Dpn I
Alw I
Fok I
Mnl I
HinP I
Hha I
BstU I
Dpn I
Nla III
Gsu I
Bpm I
Mbo II
Alu I
Pnu4H I
Bbv I
GACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTA 4400
CTGCGGCCGACCTACTAGGAGGTCGCGCCCCTAGAGTACGACCTCAAGAAGCGGGTGGGGTTGAACAAATAACGTCGAAT
4321
4321
4321
4324
4324
4324
4324
4324
4324
4325
4325
4331
4335
4335
4335
4335
4339
4341
4338
4344
4344
4345
4350
4350
4351
4351
4351
4356
4360
4360
4367
4393
4393
4395

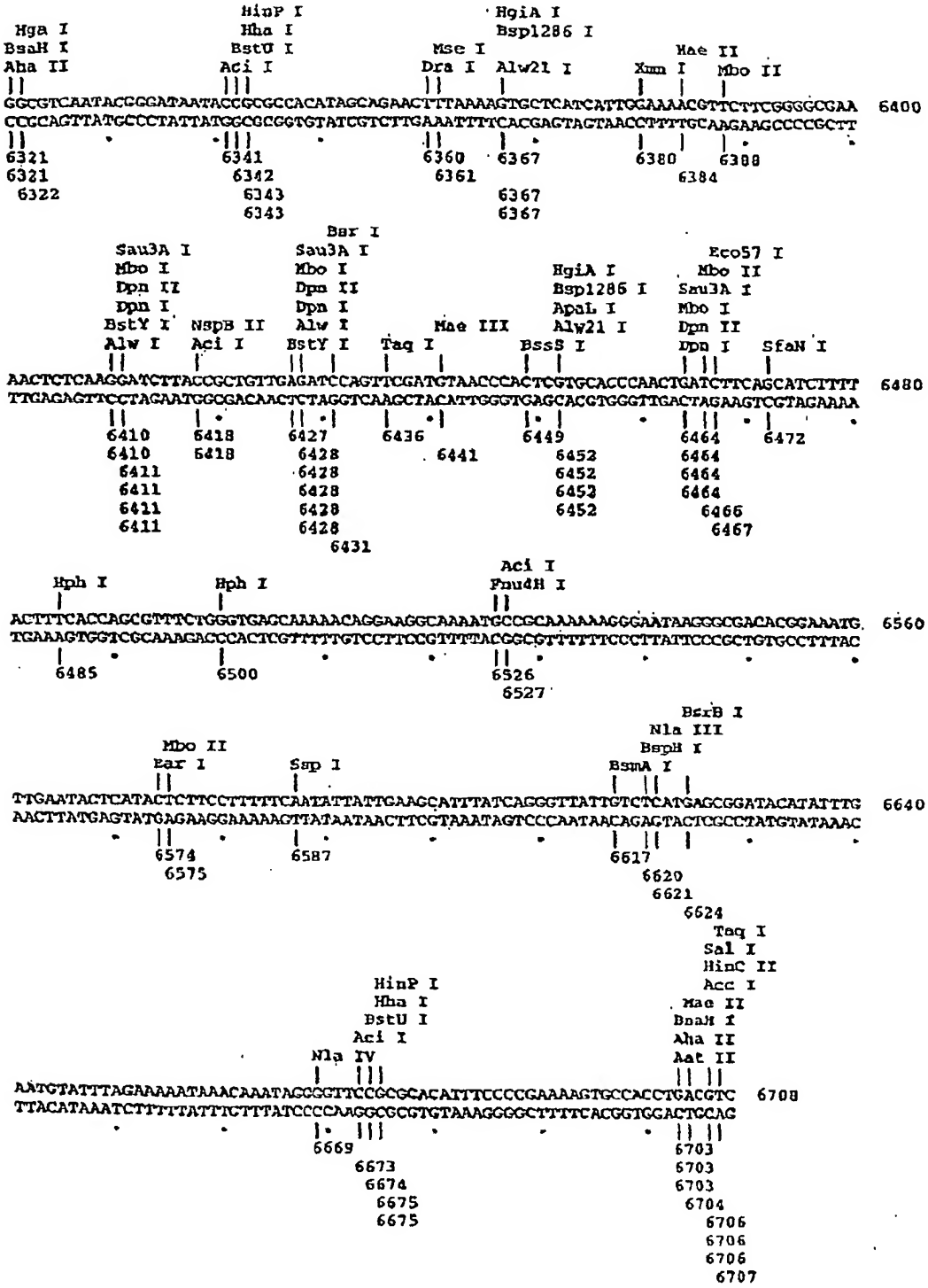

Rma I
Mae I
Bfa I
Mae III
SfaN I
Bsm I
TAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGT 4480
ATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACA
4406
4424
4461
4466
4466
4466
Rma I
Mae I
Taq I
Rma I
Xca I
Sal I
Mae I
BstI107 I
Mnl I
Bfa I
Acc I
HinC II
Alu I
Nla III
Acc I
Bfa I
Alu I
Nla III
Alu I
CCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTG 4560
GGTTTGAGTAGTTACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACCGCATTAGTACCAGTATCGAC
4504
4511
4511
4511
4518
4518
4519
4523
4526
4526
4528
4530
4530
4530
4534
4547
4556
Nla IV
Ban I
Sec I
ScrF I
EcoR II
BstN I
BsaJ I
Msp I
Hpa II
Aci I
Apo I
TTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCC 4640
AAAGGACACACTTTAACAATAGGCGAGTGTTAAGGTGTGTTGTATGCTCGGCCTTCGTATTTCACATTTCGGACCCCACG
4582
4589
4610
4610
4631
4631
4631
4631
4631
4636
4636
Fnu4H I
Bbv I
Alu I
Pvu II
NspB II
Cac8 I
Mse I
HinP I
Aci I
Alu I
Ase I
Hha I
Cac8 I
Bsr I
CTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGC 4720
GATTACTCACTCGATTGAGTGTAATTAACGCAACGCGAGTGACGGGCGAAAGGTCAGCCCTTTGGACAGCACGGTCGACG
4651
4662
4663
4674
4674
4683
4685
4692
4712
4714
4714
4715
4716
4716


BstU I
Hae III
Tfi I Bsl I
Hinf I
HinP I
Hha I
Nse I
Gdi II
Ase I
Eae I
BstU I
Mnl I
Mbo II
Ear I
HinP I
Hha I
Hae II
Aci I
Mnl I
Fnu4H I
Bbv I
Ple I
Hinf I
HinP I
Hha I
ATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTG 4800
TAATTACTTAGCCGGTTGCGCGCCCCTCTCCGCCAAACGCATAACCCGCGAGAAGGCGAAGGAGCGAGTGACTGAGCGAC
4721 4731 4738 4748 4766 4775 4792 4800
4722 4731 4739 4767 4781 4792 4800
4727 4739 4767 4797
4727 4734 4770 4797
4732 4771
4740
Cac8 I
Fnu4H I
Fnu4H I
BsiE I
Bbv I
BsrB I
Alu I
Tfi I
Hinf I
CGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAA 4880
GCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTATGCCAATAGGTGTCTTAGTCCCCTATT
4805 4815 4823 4833 4867
4815 4867
4818
4821
Nla IV
Nla III
NspH I
Nsp7524 I
Nsp I
Afl III
Bsl I
Cac8 I
Hae III
Hae I
ScrF I
EcoR II
BstN I
Hae III
Hae I
Bsl I
BstU I
Aci I
Fnu4H I
Hae III
Cac8 I
CGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA 4960
GCGTCCTTTCTTGTACACTCGTTTTCCGGTCGTTTTCCGGTCCTTCGCATTTTTCCGGCGCAACGACCGCAAAAAGGTAT
4892 4905 4916 4926 4935 4944
4892 4906 4917 4936
4892 4907 4919 4937
4892 4908 4919 4938
4893 4919
4922
Aci I
Nla IV
SfaN I
Hga I
Drd I
Taq I
Mnl I
ScrF I
EcoR II
BstN I
GGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAC 5040
CCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATATTTCTATG
4961 4980 4992 5007 5040
4965 4994 5040
4994 5040
ScrF I
EcoR II
BstN I
Sec I
BsaJ I
Alu I
BssS I
Mnl I
HinP I
Hha I
Mme I
Bsl I
Aci I
DsaW I
Msp I
Hpa II
Fnu4H I
Aci I
CAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCT 5120
GTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGA
5052 5064 5082 5091 5111
5052 5059 5070 5092 5099
5053 5065 5092 5099
5053 5070 5098
5053
HinP I
Hha I
Hae II
Sfc I
Dde I
Alu I
CCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGG 5200
GGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGACC
5136 5157 5167 5195
5137
5137


HinP I
Fnu4H I
Msp I
Hpa II
HgiA I
Bbv I
Mae III
SerF I
Bsp1286 I
NspB II
Msp I
Mme I
ApaL I
Aci I
Hpa II
Ple I
Nci I
Alw21 I
BsiE I
Hha I
BsaW I
Hinf I
Bcn I
GCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGA 5280
CGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCT
5206 5229 5237 5245 5263 5271
5206 5232 5246 5263 5271
5206 5232 5246 5266
5206 5234 5249 5271
5234 5272
5237 5272
AlwN I
Fnu4H I
Pnu4H I
Bbv I
Bsr I
Bsr I
Bbv I
Mae III
Mnl I
Sfc I
CACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT 5360
GTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAA
5295 5302 5312 5331 5348
5299 5308
5299
5302
5303
Bsl I
Rma I
Hae III
Mae I
HinP I
Bsr I
Hae I
Bfa I
Hha I
Eco57 I
Mae III
GAAGTGGTCGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAA 5440
CTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTT
5368 5387 5411 5419 5428
5369 5387 5411 5425
5371 5387
Msp I
Hpa II
Sau3A I
Fnu4H I
Mbo I
Bbv I
Dpn II
NspB II
Tth111 II
HinP I
Dpn I
HgiE II
Cac8 I
Hha I
Alu I
Alw I
Tth111 II
Aci I
Tth111 II
BstU I
AAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACG 5520
TTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGC
5452 5459 5466 5477 5499 5519
5459 5473 5504 5520
5459 5477 5505 5520
5459 5508
5459 5508
5462
5462
Sau3A I
Sau3A I
Mbo I
Mbo I
Dpn II
Dpn I
Sau3A I
Alw I
Mbo I
BstY I
BstY I
Dpn II
Dde I
Mse I
Alw I
Mbo II
Dpn I
Hga I
Mae II
CGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTA 5600
GCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAAT
5533 5542 5553 5572 5595
5533 5544 5553 5576 5598
5534 5545 5553
5534 5545 5553
5534 5545
5534 5545
5545


Sau3A I
Hph I
Mbo I
Mbo II
Dpn II
Sau3A I
Dpn I
Mbo I
Alw I
Dpn II
BstY I
Dpn I
Rma I
Nla III
Mse I
BstY I
Mae I
Mse I
BspH I
Alw I
Bfa I
Dra I
Mse I
Dra I
AGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCT 5680
TCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGA
5612
5613
5630
5630
5631
5631
5631
5633
5636
5640
5640
5640
5642
5643
5643
5643
5643
5643
5649
5650
5655
5668
5669
Sau3A I
Mbo I
Nla IV
Dpn II
Ban I
Dpn I
Mae III
Mse I
Mnl I
Dde I
AAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTT 5760
TTTCATATATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAA
5711
5722
5731
5733
5733
5742
5748
5748
5748
5748
Bsr I
Ple I
Sau96 I
BsrD I
Hinf I
Nla IV
Fnu4H I
Pok I
Eam1105 I
Mnl I
Hae III
Bbv I
CGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGC 5840
GCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACG
5765
5780
5780
5780
5812
5827
5827
5827
5831
5836
5836
5839
Nla IV
Msp I
BsmA I
Hpa II
Msp I
HinP I
Bsa I
Cfr10 I
Hpa II
Hha I
BstU I
BsrF I
Bgl I
Hae III
Aci I
Nph I
Gsu I
CacB I
Sau96 I
AATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAA 5920
TTACTATGGCGCTCTGGGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTT
5840
5848
5849
5852
5852
5863
5865
5865
5866
5866
5868
5870
5895
5899
5900
5900
5906
5907
5913
5913


Alu I
ScrF I
Nci I
Msp I
Hpa II
Bcn I
Rma I
Mae I
Bfa I
Sau96 I
Ava II
Mnl I
Aci I
Bsr I
Pok I
Mse I
Ase I
Bsr I
GTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAAT 6000
CACCAGGACGTTGAAATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTA
5923
5923
5939
5942
5946
5949
5956
5957
5967
5967
5967
5967
5967
5973
5975
5975
5992
5996
Mae II
HinF I
Hha I
Fsp I
BsrD I
Sfc I
SfaN I
Mae III
Nla IV
Msp I
Hpa II
BsaW I
Alu I
AGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGG 6080
TCAAACGCGTTGCAACAACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCC
6005
6006
6006
6011
6021
6025
6032
6042
6073
6076
6077
6077
6079
Sau3A I
Mbo I
Dpn II
Dpn I
Alw I
Nla III
Mae III
Nla III
Alu I
Sau96 I
Ava II
Dpn I
Pvu I
BsiE I
Mnl I
TTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTG 6160
AAGGGTTGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAAC
6089
6089
6089
6089
6100
6104
6107
6107
6107
6107
6107
6114
6136
6145
6145
6148
6152
6152
6153
6153
6153
Aci I
Fnu4H I
Hae III
Cdi II
Eae I
Nla III
Fnu4K I
Bbv I
Apo I
Nla III
Fok I
TCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTA 6240
AGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCAT
6173
6173
6174
6175
6176
6192
6202
6202
6213
6228
6233
Hph I
Bsr I
Rsa I
SfaN I
Mae III
Sca I
Dde I
Fnu4H I
BsiE I
Msp I
Hpa II
ScrF I
Nci I
Bcn I
AGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCC 6320
TCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGG
6242
6253
6256
6259
6263
6264
6282
6297
6301
6318
6318
6318
6319
6319



Restriction Endonucleases site usage

Aat II	6	Bsm I	1	Fnu4H I	33	Nsp7524 I	6
Acc I	7	BsmA I	10	Fok I	12	NspB II	8
Acc65 I	1	Bsp1286 I	11	Psp I	1	NspH I	6

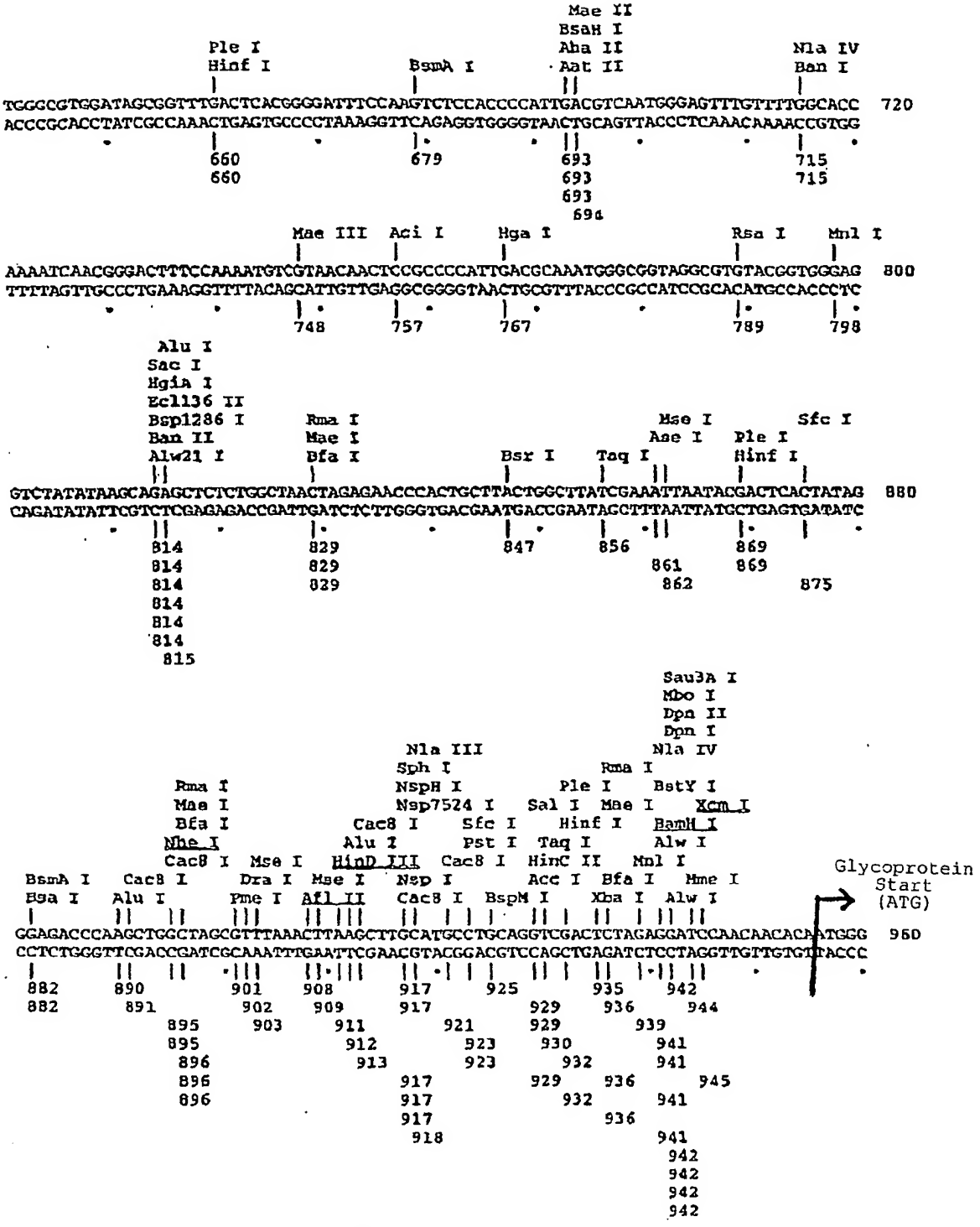
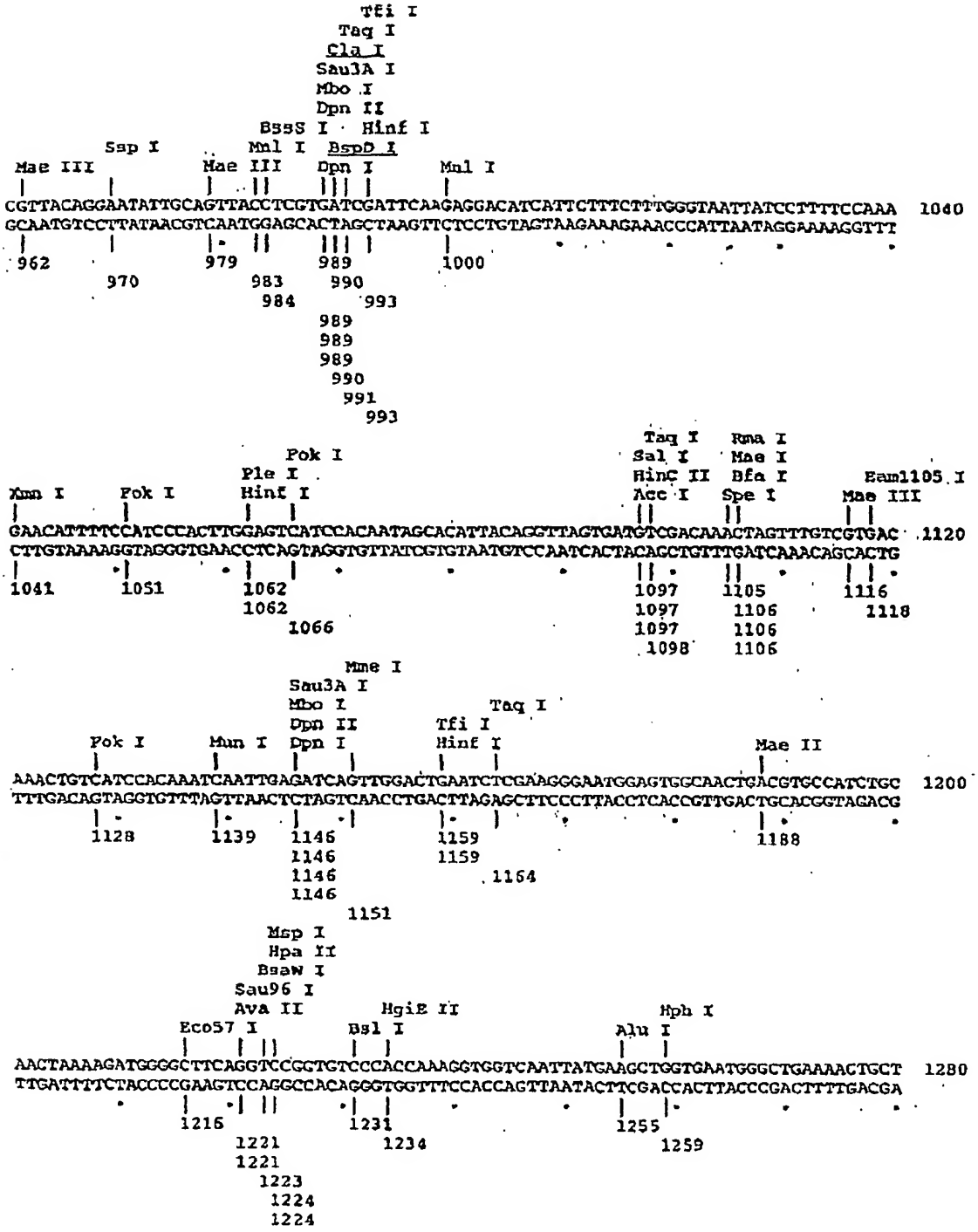
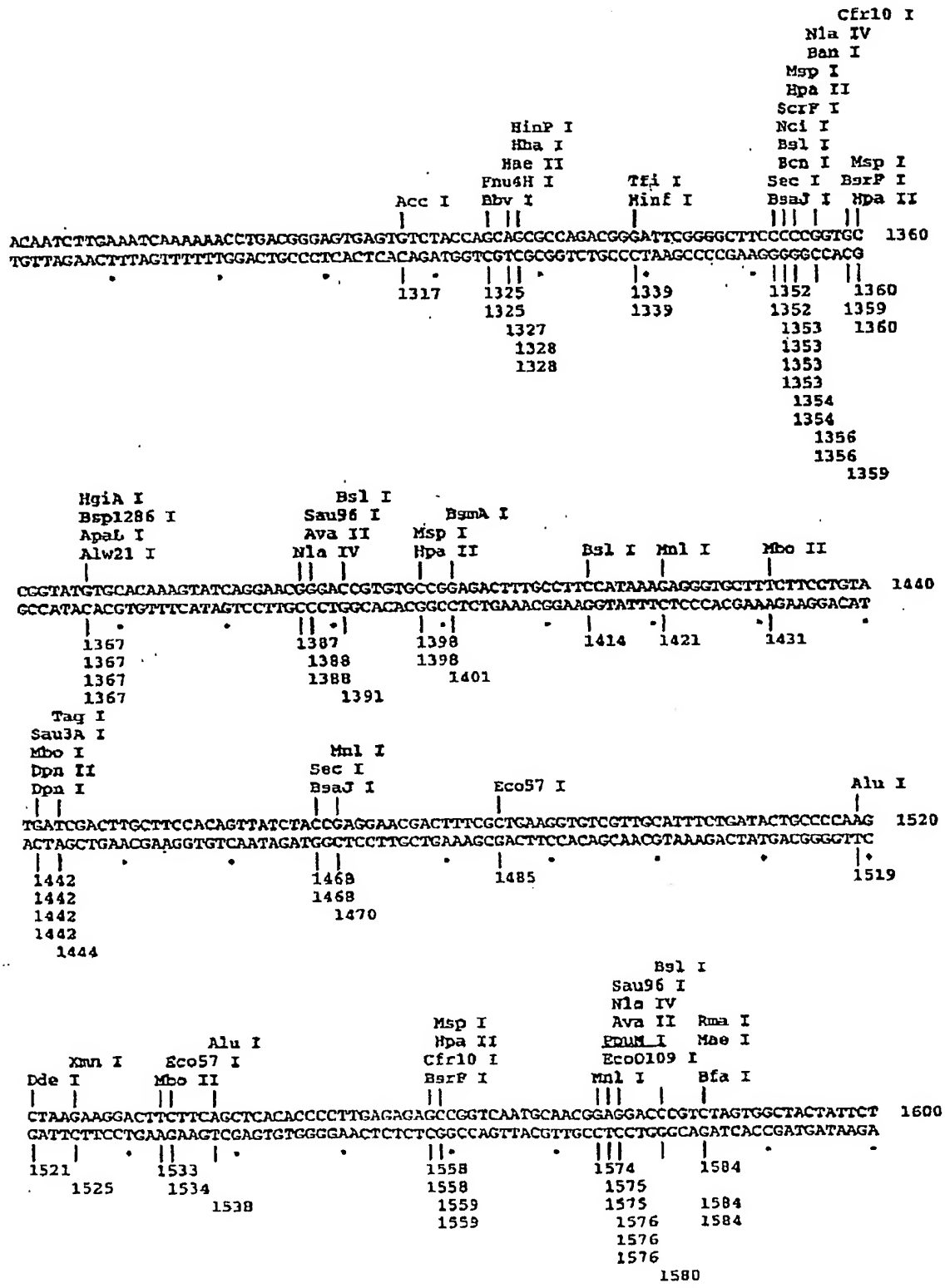
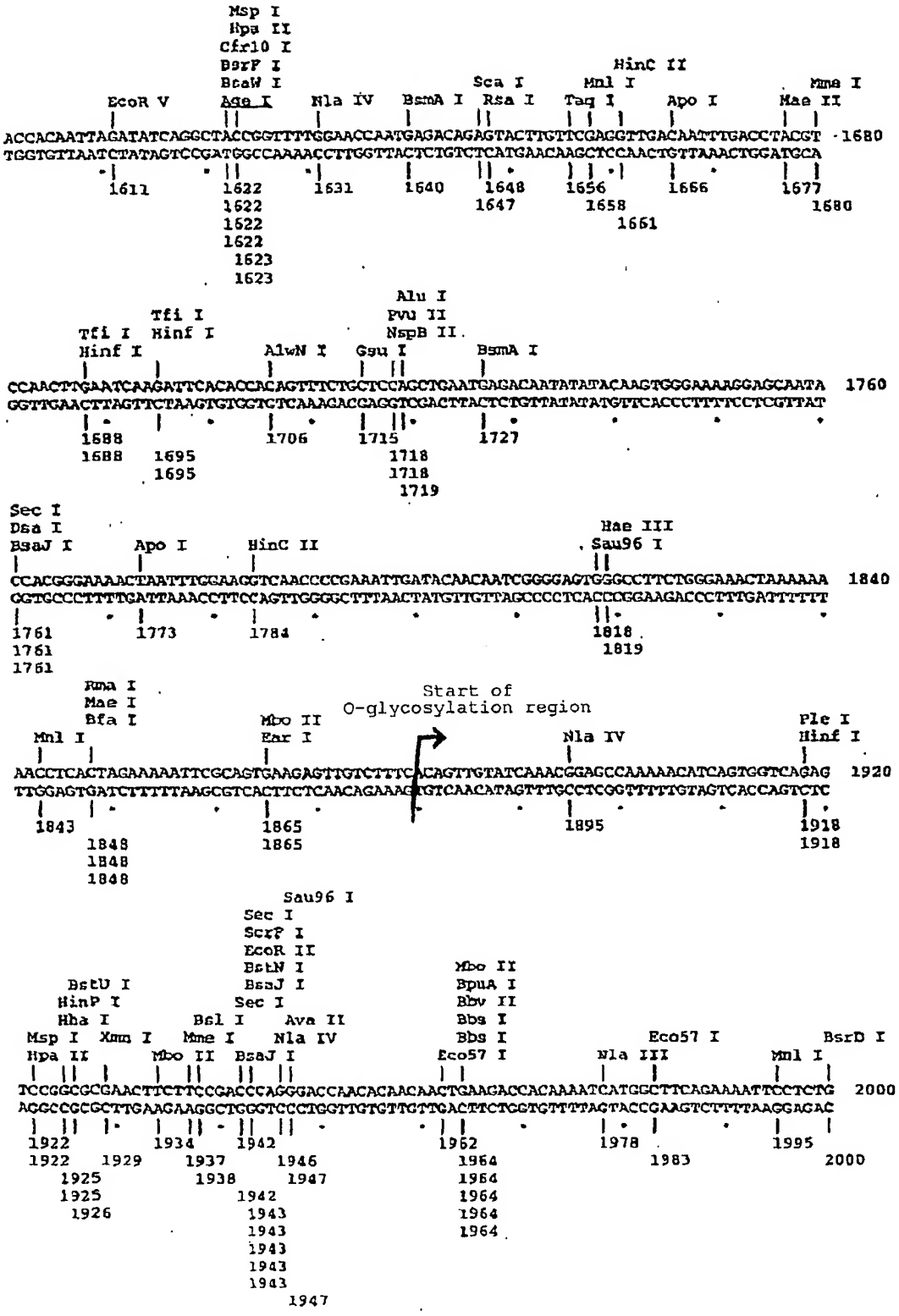
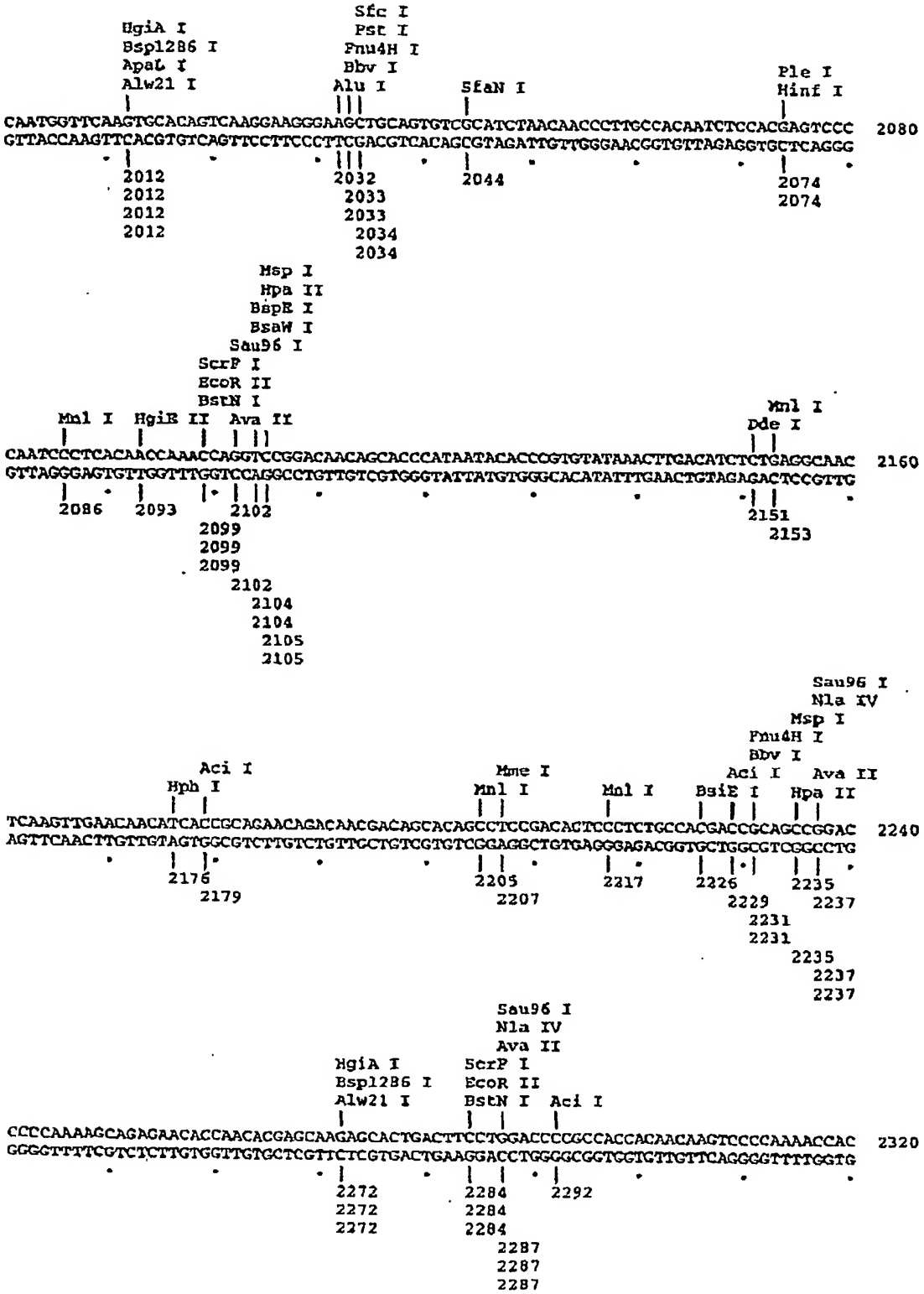

Mae II
BsaH I
Aha II
Aat II
Ple I
Hinf I
BsmA I
Nla IV
Ban I
TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC 720
ACCCGCACCTATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACCGTGG
660 679 693 715
660 693 715
693
694
Mae III
Aci I
Hga I
Rsa I
Mnl I
AAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAG 800
TTTTAGTTGCCCTGAAAGGTTTTACAGCATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCCACCCTC
748 757 767 789 798
Alu I
Sac I
HgiA I
Ecl136 II
Bsp1286 I
Ban II
Alw21 I
Rma I
Mae I
Bfa I
Bsr I
Taq I
Mse I
Ase I
Ple I
Hinf I
Sfc I
GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAG 880
CAGATATATTCGTCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATC
814 829 847 856 861 869 875
814 829 862 869
814 829
814
814
814
815
Sau3A I
Mbo I
Dpn II
Dpn I
Nla III
Nla IV
Sph I
Rma I
NspH I
Ple I
BstY I
Rma I
Mae I
Nsp7524 I
Sal I
Mae I
Xcm I
Bfa I
Cac8 I
Sfc I
Hinf I
BamH I
Nhe I
Alu I
Pst I
Taq I
Alw I
Cac8 I
Mse I
HinD III
Cac8 I
HinC II
Mnl I
BsmA I
Cac8 I
Dra I
Mse I
Nsp I
Acc I
Bfa I
Mme I
Bga I
Alu I
Pme I
Afl II
Cac8 I
BspM I
Xba I
Alw I
Glycoprotein Start (ATG)
GGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCAACAACACAATGGG 960
CCTCTGGGTTCGACCGATCGCAAATTTGAATTCGAACGTACGGACGTCCAGCTGAGATCTCCTAGGTTGTTGTGTTACCC
882 890 901 908 917 925 935 942
882 891 902 909 917 929 936 944
895 903 911 921 929 939
895 912 923 930 941
896 913 923 932 941
896 917 929 936 945
896 917 932 941
917 936
918 941
942
942
942
942

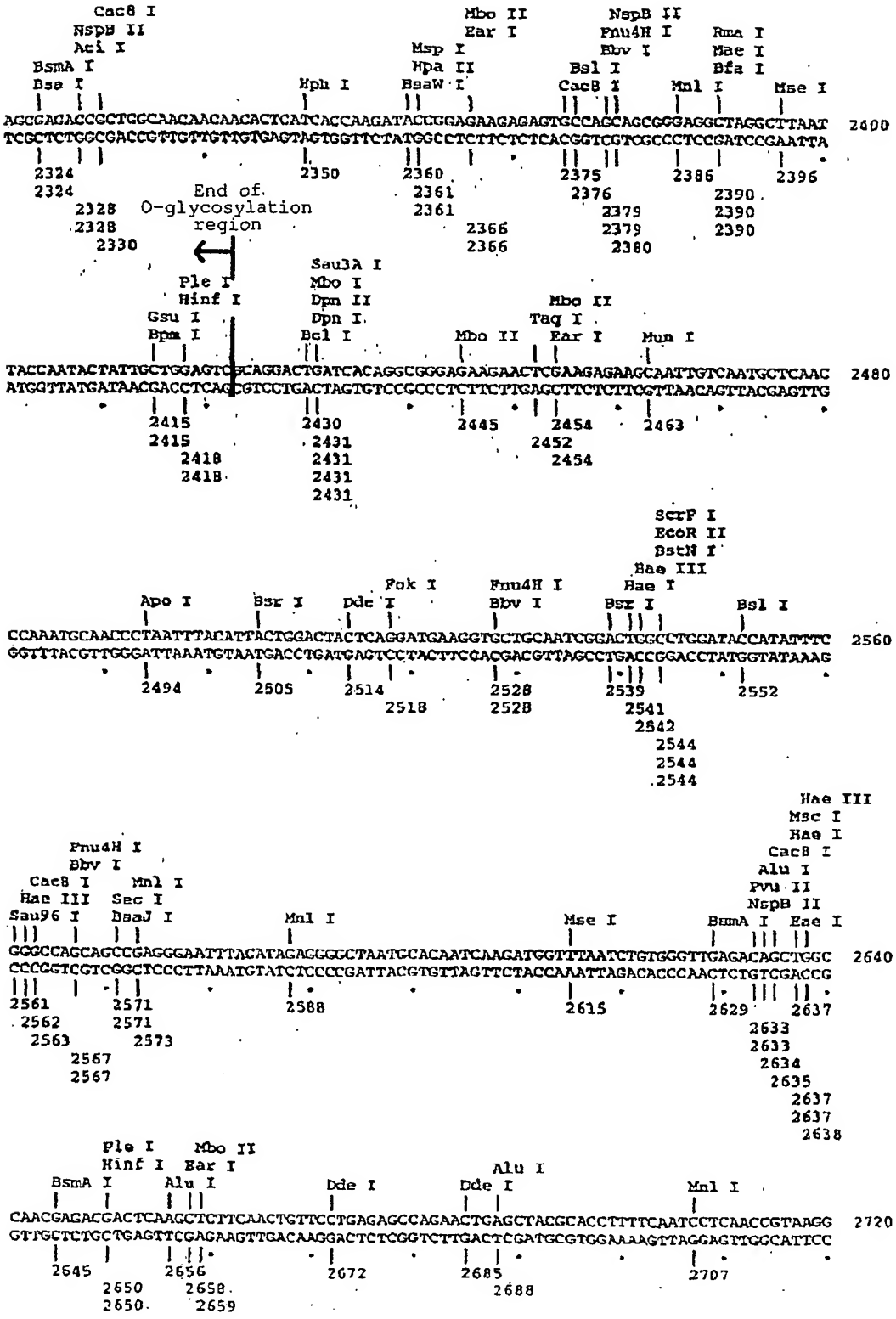
Fig. 1C


Tfi I
Taq I
Cla I
Sau3A I
Mbo I
Dpn II
BssS I
Hinf I
Ssp I
Mnl I
BspD I
Mae III
Mae III
Dpn I
Mnl I
CGTTACAGGAATATTGCAGTTACCTCGTGATCGATTCAAGAGGACATCATTCTTTCTTTGGGTAATTATCCTTTTCCAAA 1040
GCAATGTCCTTATAACGTCAATGGAGCACTAGCTAAGTTCTCCTGTAGTAAGAAAGAAACCCATTAATAGGAAAAGGTTT
962
970
979
983
984
989
990
993
989
989
989
990
991
993
Taq I
Rma I
Pok I
Sal I
Mae I
Ple I
Hinc II
Bfa I
Eam1105 I
Xmn I
Pok I
Hinf I
Acc I
Spe I
Mae III
GAACATTTTCCATCCCACTTGGAGTCATCCACAATAGCACATTACAGGTTAGTGATGTCGACAAACTAGTTTGTCGTGAC 1120
CTTGTAAAAGGTAGGGTGAACCTCAGTAGGTGTTATCGTGTAATGTCCAATCACTACAGCTGTTTGATCAAACAGCACTG
1041
1051
1062
1062
1066
1097
1097
1097
1098
1105
1106
1106
1106
1116
1118
Mme I
Sau3A I
Mbo I
Taq I
Dpn II
Tfi I
Pok I
Mun I
Dpn I
Hinf I
Mae II
AAACTGTCATCCACAAATCAATTGAGATCAGTTGGACTGAATCTCGAAGGGAATGGAGTGGCAACTGACGTGCCATCTGC 1200
TTTGACAGTAGGTGTTTAGTTAACTCTAGTCAACCTGACTTAGAGCTTCCCTTACCTCACCGTTGACTGCACGGTAGACG
1128
1139
1146
1146
1146
1146
1151
1159
1159
1164
1188
Msp I
Hpa II
BsaW I
Sau96 I
Ava II
HgiE II
Hph I
Eco57 I
Bsl I
Alu I
AACTAAAAGATGGGGCTTCAGGTCCGGTGTCCCACCAAAGGTGGTCAATTATGAAGCTGGTGAATGGGCTGAAAACTGCT 1280
TTGATTTTCTACCCCGAAGTCCAGGCCACAGGGTGGTTTCCACCAGTTAATACTTCGACCACTTACCCGACTTTTGACGA
1216
1221
1221
1223
1224
1224
1231
1234
1255
1259


Cfr10 I
Nla IV
Ban I
Msp I
Hpa II
ScrF I
Nci I
Bsl I
Bcn I Msp I
Sec I BsrF I
BsaJ I Hpa II
HinP I
Hha I
Hae II
Fnu4H I
Bbv I
Acc I
Tfi I
Hinf I
ACAATCTTGAAATCAAAAAACCTGACGGGAGTGAGTGTCTACCAGCAGCGCCAGACGGGATTCGGGGCTTCCCCCGGTGC 1360
TGTTAGAACTTTAGTTTTTTGGACTGCCCTCACTCACAGATGGTCGTCGCGGTCTGCCCTAAGCCCCGAAGGGGGCCACG
1317 1325 1339 1352 1360
1325 1339 1352 1359
1327 1353 1360
1328 1353
1328 1353
1353
1354
1354
1356
1356
1359
HgiA I
Bsp1286 I
ApaL I
Alw21 I
Bsl I
Sau96 I
Ava II
Nla IV
BsmA I
Msp I
Hpa II
Bsl I Mnl I Mbo II
CGGTATGTGCACAAAGTATCAGGAACGGGACCGTGTGCCGGAGACTTTGCCTTCCATAAAGAGGGTGCTTTCTTCCTGTA 1440
GCCATACACGTGTTTCATAGTCCTTGCCCTGGCACACGGCCTCTGAAACGGAAGGTATTTCTCCCACGAAAGAAGGACAT
1367 1387 1398 1414 1421 1431
1367 1388 1398
1367 1388 1401
1367 1391
Tag I
Sau3A I
Mbo I
Dpn II
Dpn I
Mnl I
Sec I
BsaJ I
Eco57 I
Alu I
TGATCGACTTGCTTCCACAGTTATCTACCGAGGAACGACTTTCGCTGAAGGTGTCGTTGCATTTCTGATACTGCCCCAAG 1520
ACTAGCTGAACGAAGGTGTCAATAGATGGCTCCTTGCTGAAAGCGACTTCCACAGCAACGTAAAGACTATGACGGGGTTC
1442 1468 1485 1519
1442 1468
1442 1470
1442
1444
Bsl I
Sau96 I
Nla IV
Msp I
Hpa II
Ava II
PpuM I
Rma I
Mae I
Alu I
Xmn I
Eco57 I
Cfr10 I
Eco0109 I
Dde I
Mbo II
BsrF I
Mnl I
Bfa I
CTAAGAAGGACTTCTTCAGCTCACACCCCTTGAGAGAGCCGGTCAATGCAACGGAGGACCCGTCTAGTGGCTACTATTCT 1600
GATTCTTCCTGAAGAAGTCGAGTGTGGGGAACTCTCTCGGCCAGTTACGTTGCCTCCTGGGCAGATCACCGATGATAAGA
1521 1533 1558 1574 1584
1525 1534 1558 1575
1538 1559 1575 1584
1559 1576 1584
1576
1576
1580


Msp I
Hpa II
Cfr10 I
BsrF I
BstaW I
Age I
EcoR V
Nla IV
BsmA I
Sca I
Rsa I
HinC II
Mnl I
Taq I
Apo I
Mae II
Mme I
ACCACAATTAGATATCAGGCTACCGGTTTTGGAACCAATGAGACAGAGTACTTGTTCGAGGTTGACAATTTGACCTACGT 1680
TGGTGTTAATCTATAGTCCGATGGCCAAAACCTTGGTTACTCTGTCTCATGAACAAGCTCCAACTGTTAAACTGGATGCA
1611
1622
1622
1622
1622
1623
1623
1631
1640
1647
1648
1656
1658
1661
1666
1677
1680
Alu I
Pvu II
NspB II
Tfi I
Hinf I
AlwN I
Gsu I
BsmA I
CCAACTTGAATCAAGATTCACACCACAGTTTCTGCTCCAGCTGAATGAGACAATATATACAAGTGGGAAAAGGAGCAATA 1760
GGTTGAACTTAGTTCTAAGTGTGGTGTCAAAGACGAGGTCGACTTACTCTGTTATATATGTTCACCCTTTTCCTCGTTAT
1688
1688
1695
1695
1706
1715
1718
1718
1719
1727
Sec I
Dsa I
BsaJ I
Apo I
HinC II
Hae III
Sau96 I
CCACGGGAAAACTAATTTCGAAGGTCAACCCCGAAATTGATACAACAATCGGGGAGTGGGCCTTCTGGGAAACTAAAAAA 1840
GGTGCCCTTTTGATTAAACCTTCCAGTTGGGGCTTTAACTATGTTGTTAGCCCCTCACCCGGAAGACCCTTTGATTTTTT
1761
1761
1761
1773
1784
1818
1819
Rma I
Mae I
Bfa I
Mnl I
Mbo II
Ear I
Start of
O-glycosylation region
Nla IV
Ple I
Hinf I
AACCTCACTAGAAAAATTCGCAGTGAAGAGTTGTCTTTCACAGTTGTATCAAACGGAGCCAAAAACATCAGTGGTCAGAG 1920
TTGGAGTGATCTTTTTAAGCGTCACTTCTCAACAGAAAGTGTCAACATAGTTTGCCTCGGTTTTTGTAGTCACCAGTCTC
1843
1848
1848
1848
1865
1865
1895
1918
1918
Sau96 I
Sec I
ScrF I
EcoR II
BstN I
BsaJ I
Sec I
Mbo II
BpuA I
Bbv II
Bbs I
Bbs I
Eco57 I
BstU I
HinP I
Hha I
Bsl I
Ava II
Msp I
Xmn I
Mme I
Nla IV
Hpa II
Mbo II
BsaJ I
Nla III
Eco57 I
Mnl I
BsrD I
TCCGGCGCGAACTTCTTCCGACCCAGGGACCAACACAACAACTGAAGACCACAAAATCATGGCTTCAGAAAATTCCTCTG 2000
AGGCCGCGCTTGAAGAAGGCTGGGTCCCTGGTTGTGTTGTTGACTTCTCGTGTTTTAGTACCGAAGTCTTTTAAGGAGAC
1922
1922
1925
1925
1926
1929
1934
1937
1938
1942
1942
1943
1943
1943
1943
1943
1946
1947
1947
1962
1964
1964
1964
1964
1964
1978
1983
1995
2000


Sfc I
Pst I
Fnu4H I
Bbv I
HgiA I
Bsp1286 I
ApaL I
Alw21 I
Alu I
SfaN I
Ple I
Hinf I
CAATGGTTCAAGTGCACAGTCAAGGAAGGGAAGCTGCAGTGTCGCATCTAACAACCCTTGCCACAATCTCCACGAGTCCC 2080
GTTACCAAGTTCACGTGTCAGTTCCTTCCCTTCGACGTCACAGCGTAGATTGTTGGGAACGGTGTTAGAGGTGCTCAGGG
2012
2012
2012
2012
2032
2033
2033
2034
2034
2044
2074
2074
Msp I
Hpa II
BspR I
BsaW I
Sau96 I
ScrF I
EcoR II
BstN I
Mnl I
HgiE II
Ava II
Dde I
CAATCCCTCACAACCAAACCAGGTCCGGACAACAGCACCCATAATACACCCGTGTATAAACTTGACATCTCTGAGGCAAC 2160
GTTAGGGAGTGTTGGTTTGGTCCAGGCCTGTTGTCGTGGGTATTATGTGGGCACATATTTGAACTGTAGAGACTCCGTTG
2086
2093
2102
2099
2099
2099
2102
2104
2104
2105
2105
2151
2153
Sau96 I
Nla IV
Msp I
Fnu4H I
Bbv I
Aci I
Hph I
Mme I
Mnl I
Mnl I
BsiE I
Ava II
Hpa II
TCAAGTTGAACAACATCACCGCAGAACAGACAACGACAGCACAGCCTCCGACAGTCCCTCTGCCACGACCGCAGCCGGAC 2240
AGTTCAACTTGTTGTAGTGGCGTCTTGTCTGTTGCTGTCGTGTCGGAGGCTGTGAGGGAGACGGTGCTGGCGTCGGCCTG
2176
2179
2205
2207
2217
2226
2229
2231
2231
2235
2235
2237
2237
Sau96 I
Nla IV
Ava II
HgiA I
Bsp1286 I
Alw21 I
ScrF I
EcoR II
BstN I
Aci I
CCCCAAAAGCAGAGAACACCAACACGAGCAAGAGCACTGACTTCCTGGACCCCGCCACCACAACAAGTCCCCAAAACCAC 2320
GGGGTTTTCGTCTCTTGTGGTTGTGCTCGTTCTCGTGACTGAAGGACCTGGGGCGGTGGTGTTGTTCAGGGGTTTTGGTG
2272
2272
2272
2284
2284
2284
2287
2287
2287
2292


Cac8 I
NspB II
Aci I
BsmA I
Bse I
Hph I
Msp I
Hpa II
BsaW I
Mbo II
Ear I
NspB II
Fnu4H I
Bbv I
Bsl I
CacB I
Rma I
Mae I
Bfa I
Mnl I
Mse I
AGCGAGACCGCTGGCAACAACAACACTCATCACCAAGATACCGGAGAAGAGAGTGCCAGCAGCGGGAGGCTAGGCTTAAT 2400
TCGCTCTGGCGACCGTTGTTGTTGTGAGTAGTGGTTCTATGGCCTCTTCTCTCACGGTCGTCGCCCTCCGATCCGAATTA
2324
2324
2328
2328
2330
2350
2360
2361
2361
2366
2366
2375
2376
2379
2379
2380
2386
2390
2390
2390
2396
End of
O-glycosylation
region
Ple I
Hinf I
Sau3A I
Mbo I
Dpn II
Dpn I
Bcl I
Gsu I
Bpm I
Mbo II
Taq I
Ear I
Mun I
TACCAATACTATTGCTGGAGTCGCAGGACTGATCACAGGCGGGAGAAGAACTCGAAGAGAAGCAATTGTCAATGCTCAAC 2480
ATGGTTATGATAACGACCTCAGCGTCCTGACTAGTGTCCGCCCTCTTCTTCTTGAGCTTCTCTTCGTTAACAGTTACGAGTTG
2415
2415
2418
2418
2430
2431
2431
2431
2445
2452
2454
2454
2463
ScrF I
EcoR II
BstN I
Bae III
Hae I
Fok I
Fnu4H I
Bbv I
Bsr I
Apo I
Bsr I
Dde I
Bsr I
Bsl I
CCAAATGCAACCCTAATTTACATTACTGGACTACTCAGGATGAAGGTGCTGCAATCGGACTGGCCTGGATACCATATTTC 2560
GGTTTACGTTGGGATTAAATGTAATGACCTGATGAGTCCTACTTCCACGACGTTAGCCTGACCGGACCTATGGTATAAAG
2494
2505
2514
2518
2528
2528
2539
2541
2542
2544
2544
2544
2552
Hae III
Msc I
Hae I
CacB I
Alu I
Pvu II
NspB II
Fnu4H I
Bbv I
Cac8 I
Hae III
Mnl I
Sec I
Sau96 I
BsaJ I
Mnl I
Mse I
BsmA I
Eae I
GGGCCAGCAGCCGAGGGAATTTACATAGAGGGGCTAATGCACAATCAAGATGGTTTAATCTGTGGGTTGAGACAGCTGGC 2640
CCCGGTCGTCGGCTCCCTTAAATGTATCTCCCCGATTACGTGTTAGTTCTACCAAATTAGACACCCAACTCTGTCGACCG
2561
2562
2563
2567
2567
2571
2571
2573
2588
2615
2629
2633
2633
2634
2635
2637
2637
2638
Ple I
Hinf I
Mbo II
Ear I
Alu I
BsmA I
Alu I
Dde I
Dde I
Mnl I
CAACGAGACGACTCAAGCTCTTCAACTGTTCCTGAGAGCCAGAACTGAGCTACGCACCTTTTCAATCCTCAACCGTAAGG 2720
GTTGCTCTGCTGAGTTCGAGAAGTTGACAAGGACTCTCGGTCTTGACTCGATGCGTGGAAAAGTTAGGAGTTGGCATTCC
2645
2650
2650
2656
2658
2659
2672
2685
2688
2707

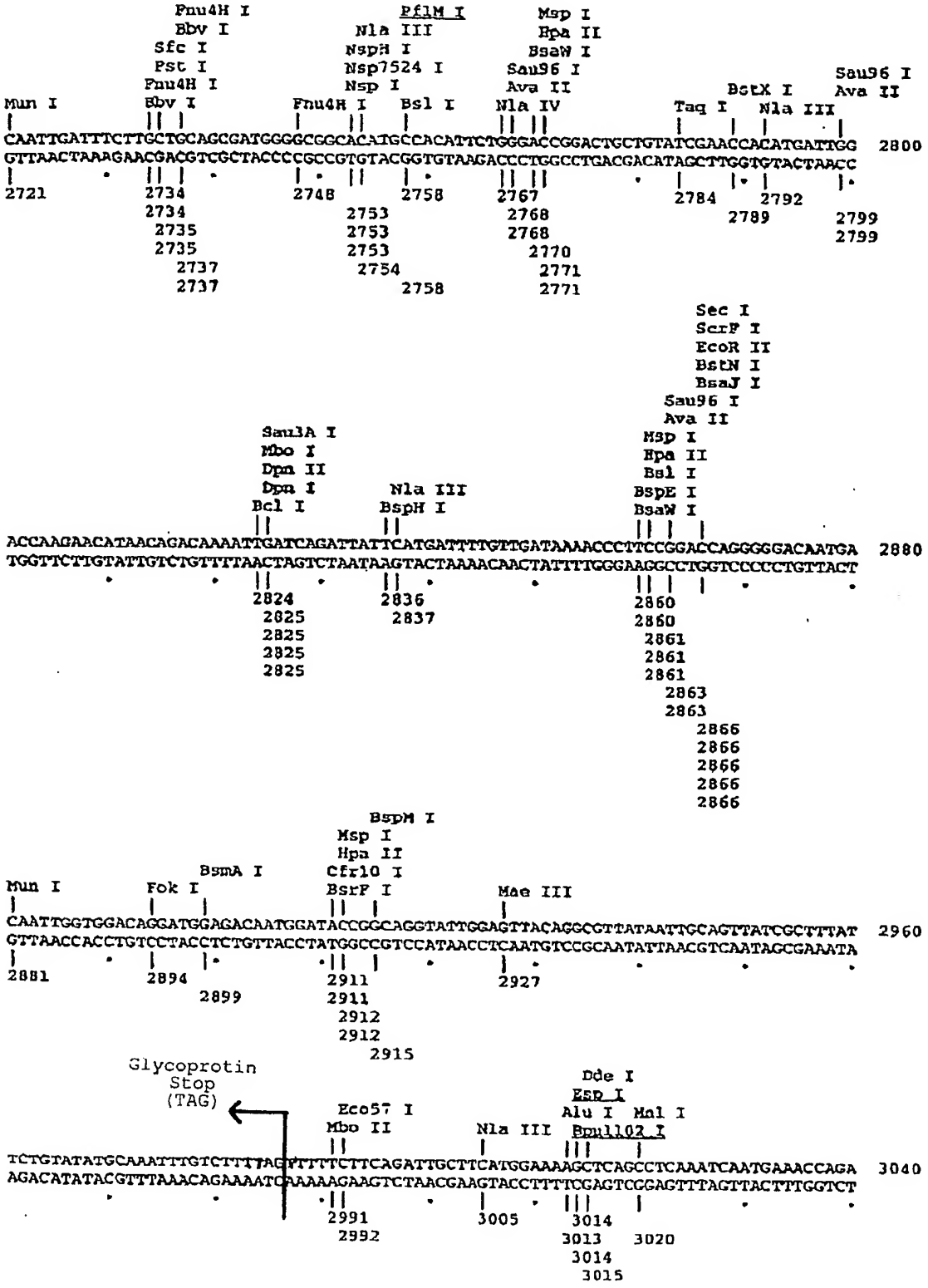

Fnu4H I
Bbv I
Sfc I
Pst I
Fnu4H I
Bbv I
Mun I
Fnu4H I
Nla III
NspH I
Nsp7524 I
Nsp I
Bsl I
PflM I
Msp I
Hpa II
BsaW I
Sau96 I
Ava II
Nla IV
Taq I
BstX I
Nla III
Sau96 I
Ava II
CAATTGATTTCTTGCTGCAGCGATGGGGCGGCACATGCCACATTCTGGGACCGGACTGCTGTATCGAACCACATGATTGG 2800
GTTAACTAAAGAACGACGTCGCTACCCCGCCGTGTACGGTGTAAGACCCTGGCCTGACGACATAGCTTGGTGTACTAACC
2721
2734
2734
2735
2735
2737
2737
2748
2753
2753
2753
2754
2758
2758
2767
2768
2768
2770
2771
2771
2784
2789
2792
2799
2799
Sec I
ScrF I
EcoR II
BstN I
BsaJ I
Sau96 I
Ava II
Sau3A I
Mbo I
Dpn II
Dpn I
Bcl I
Nla III
BspH I
Msp I
Hpa II
Bsl I
BspE I
BsaW I
ACCAAGAACATAACAGACAAAATTGATCAGATTATTCATGATTTTGTTGATAAAACCCTTCCGGACCAGGGGGACAATGA 2880
TGGTTCTTGTATTGTCTGTTTTAACTAGTCTAATAAGTACTAAAACAACTATTTTGGGAAGGCCTGGTCCCCCTGTTACT
2824
2825
2825
2825
2825
2836
2837
2860
2860
2861
2861
2861
2863
2863
2866
2866
2866
2866
2866
BspM I
Msp I
Hpa II
BsmA I
Cfr10 I
Mun I
Fok I
BsrF I
Mae III
CAATTGGTGGACAGGATGGAGACAATGGATACCGGCAGGTATTGGAGTTACAGGCGTTATAATTGCAGTTATCGCTTTAT 2960
GTTAACCACCTGTCCTACCTCTGTTACCTATGGCCGTCCATAACCTCAATGTCCGCAATATTAACGTCAATAGCGAAATA
2881
2894
2899
2911
2911
2912
2912
2915
2927
Glycoprotin
Stop
(TAG)
Eco57 I
Mbo II
Nla III
Dde I
ESP I
Alu I
Mnl I
Bpu1102 I
TCTGTATATGCAAATTTGTCTTTTAGTTTTTCTTCAGATTGCTTCATGGAAAAGCTCAGCCTCAAATCAATGAAACCAGA 3040
AGACATATACGTTTAAACAGAAAATCAAAAAGAAGTCTAACGAAGTACCTTTTCGAGTCGGAGTTTAGTTACTTTGGTCT
2991
2992
3005
3013
3014
3014
3015
3020

Fig. 2

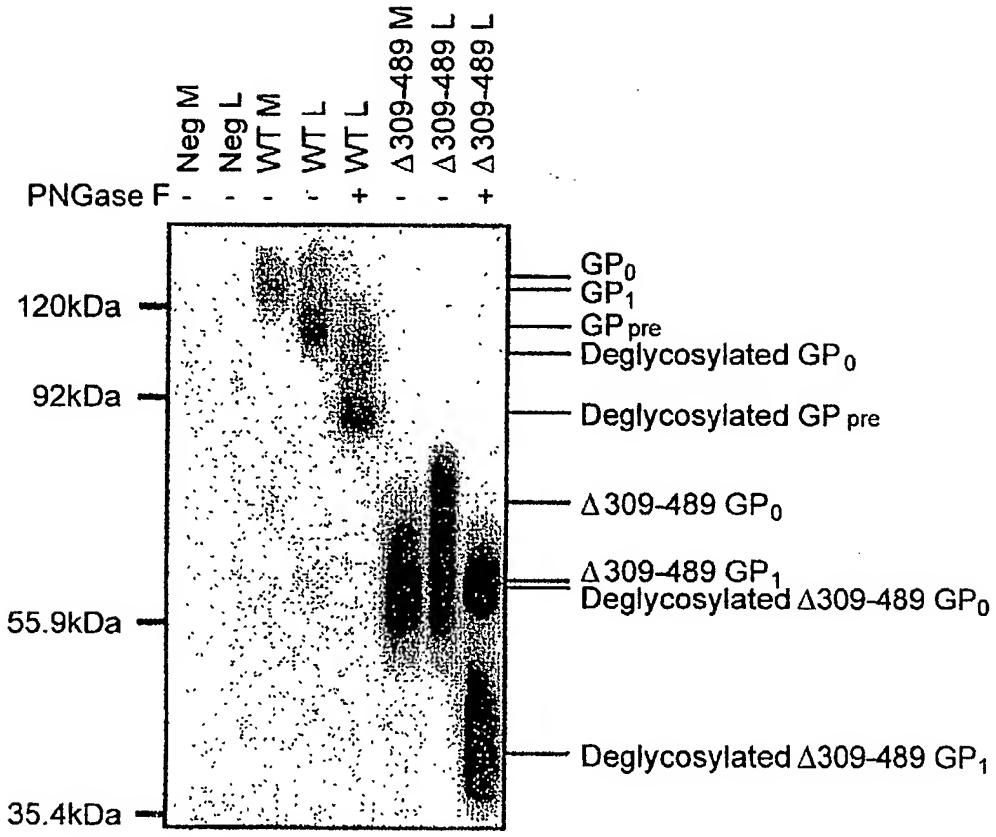
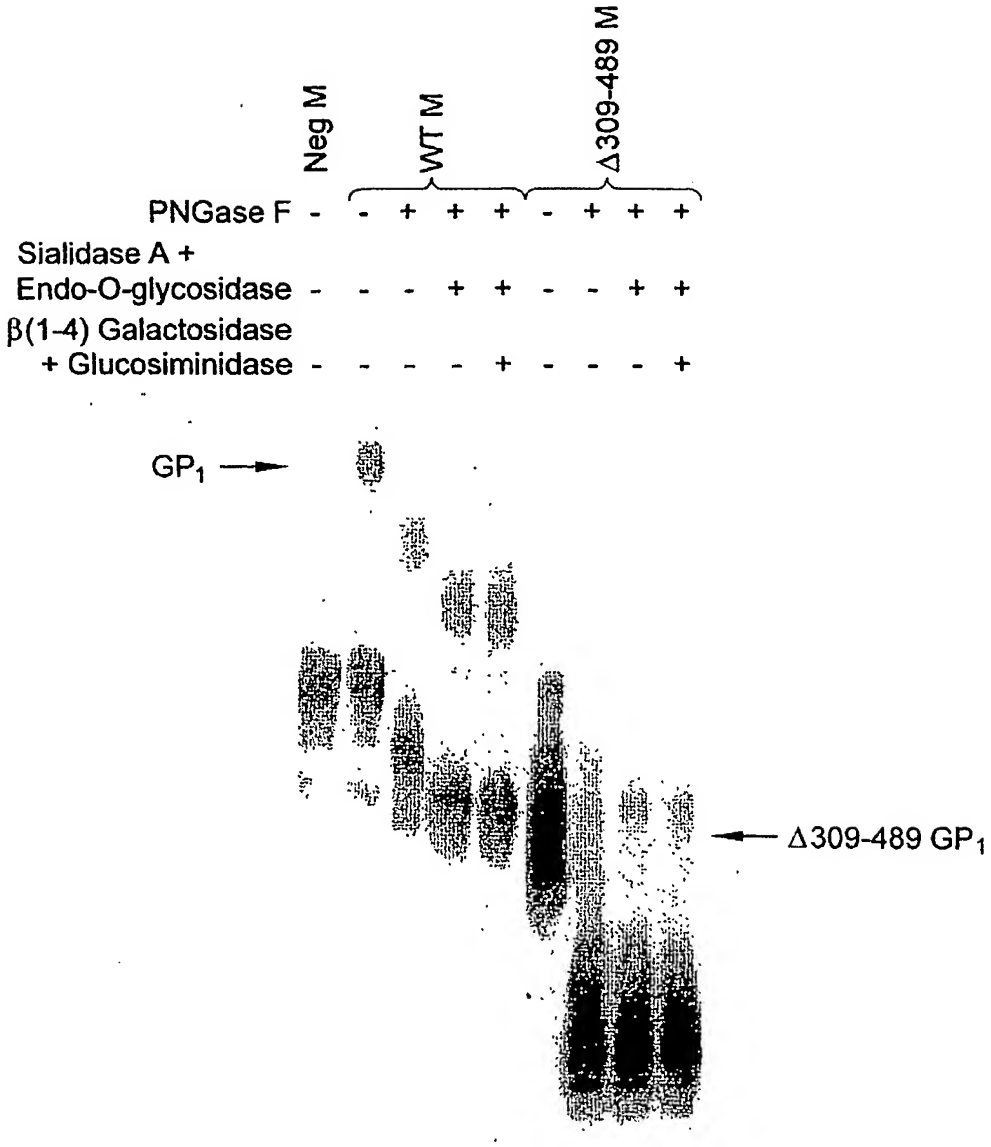

Neg M
Neg L
WT M
WT L
WT L
Δ309-489 M
Δ309-489 L
Δ309-489 L
PNGase F - - - - + - - +
120kDa
92kDa
55.9kDa
35.4kDa
GP0
GP1
GP pre
Deglycosylated GP0
Deglycosylated GP pre
Δ309-489 GP0
Δ309-489 GP1
Deglycosylated Δ309-489 GP0
Deglycosylated Δ309-489 GP1

Fig. 3


Neg M
WT M
Δ309-489 M
PNGase F - - + + + - + + +
Sialidase A +
Endo-O-glycosidase - - - + + - - + +
β(1-4) Galactosidase
+ Glucosiminidase - - - - + - - - +
GP1
Δ309-489 GP1